

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/059294 A1

(51) International Patent Classification⁷: C12N 15/09,
15/63

Bingham Circuit, Kaleen, Australian Capital Territory
2617 (AU).

(21) International Application Number: PCT/AU02/00073

(74) Agents: OLIVE, Mark, R. et al.; FB RICE & CO, 139
Rathdowne Street, Carlton, Victoria 3053 (AU).

(22) International Filing Date: 24 January 2002 (24.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/264,067 26 January 2001 (26.01.2001) US
60/333,743 29 November 2001 (29.11.2001) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, MI, MR,
NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): COM-
MONWEALTH SCIENTIFIC AND INDUSTRIAL
RESEARCH ORGANISATION [AU/AU]; Limestone
Avenue, Campbell, Australian Capital Territory 2601
(AU).

(72) Inventors; and

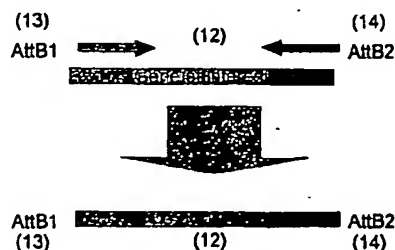
(75) Inventors/Applicants (*for US only*): WESLEY, Susan
[IN/AU]; 18 Pambula Street, Kaleen, Australian Capital
Territory 2617 (AU). WATERHOUSE, Peter [AU/AU];
5 Banjine Street, O'Connor, Australian Capital Territory
2602 (AU). HELLIWELL, Christopher [AU/AU]; 25A

Published:

— with international search report

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: METHODS AND MEANS FOR PRODUCING EFFICIENT SILENCING CONSTRUCT USING RECOMBINA-
TIONAL CLONING



(57) Abstract: Methods and means are provided for producing chimeric nu-
cleic acid constructs capable of producing dsRNA for silencing target nucleic
acid sequences of interest using recombinational cloning.

WO 02/059294 A1

BEST AVAILABLE COPY

**Methods and means for producing efficient silencing construct using
recombinational cloning.**

Field of the invention.

- 5 This invention relates to efficient methods and means for producing chimeric nucleic acid constructs capable of producing dsRNA useful for silencing target nucleic acid sequences of interest. The efficiency of the disclosed methods and means further allows high throughput analysis methods to determine the function of isolated nucleic acids, such as ESTs, without a known function and may further be put to use to
- 10 isolate particular genes or nucleotide sequences from a preselected group of genes.

General

- This specification contains nucleotide and amino acid sequence information prepared using PatentIn Version 3.1, presented herein after the claims. Each nucleotide
- 15 sequence is identified in the sequence listing by the numeric indicator <210> followed by the sequence identifier (e.g. <210>1, <210>2, <210>3, etc). The length and type of sequence (DNA, protein (PRT), etc), and source organism for each nucleotide sequence, are indicated by information provided in the numeric indicator fields <211>, <212> and <213>, respectively. Nucleotide sequences referred to in
- 20 the specification are defined by the term "SEQ ID NO:", followed by the sequence identifier (eg. SEQ ID NO: 1 refers to the sequence in the sequence listing designated as <400>1).

- The designation of nucleotide residues referred to herein are those recommended by
- 25 the IUPAC-IUB Biochemical Nomenclature Commission, wherein A represents Adenine, C represents Cytosine, G represents Guanine, T represents thymine, Y represents a pyrimidine residue, R represents a purine residue, M represents Adenine or Cytosine, K represents Guanine or Thymine, S represents Guanine or Cytosine, W represents Adenine or Thymine, H represents a nucleotide other than Guanine, B
- 30 represents a nucleotide other than Adenine, V represents a nucleotide other than Thymine, D represents a nucleotide other than Cytosine and N represents any nucleotide residue.

As used herein the term "derived from" shall be taken to indicate that a specified integer may be obtained from a particular source albeit not necessarily directly from that source.

5 Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated step or element or integer or group of steps or elements or integers but not the exclusion of any other step or element or integer or group of elements or integers.

10

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds
15 referred to or indicated in this specification, individually or collectively, and any and all combinations or any two or more of said steps or features.

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended for the purposes of exemplification only.

20 Functionally-equivalent products, compositions and methods are clearly within the scope of the invention, as described herein.

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgment or any form of suggestion that such prior art forms part of the
25 common general knowledge in Australia.

Background art.

Increasingly, the nucleotide sequence of whole genomes of organisms, including *Arabidopsis thaliana*, has been determined and as these data become available they
30 provide a wealth of unmined information. The ultimate goal of these genome projects is to identify the biological function of every gene in the genome.

Attribution of a function to a nucleic acid with a particular nucleotide sequence can be achieved in a variety of ways. Some of the genes have been characterized directly using the appropriate assays. Others have been attributed with a tentative function through homology with (parts of) genes having a known function in other organisms.

- 5 Loss-of-function mutants, obtained e.g. by tagged insertional mutagenesis have also been very informative about the role of some of these unknown genes (AzpiroLeehan and Feldmann 1997; Martienssen 1998) particularly in the large scale analysis of the yeast genome (Ross-MacDonald et al., 1999).
- 10 Structural mutants resulting in a loss-of-function may also be mimicked by interfering with the expression of a nucleic acid of interest at the transcriptional or post-transcriptional level. Silencing of genes, particularly plant genes using anti-sense or co-suppression constructs to identify gene function, especially for a larger number of targets, is however hampered by the relatively low proportion of silenced individuals
- 15 obtained, particularly those wherein the silencing level is almost complete.

Recent work has demonstrated that the silencing efficiency could be greatly improved both on quantitative and qualitative level using chimeric constructs encoding RNA capable of forming a double stranded RNA by basepairing between the antisense and

- 20 sense RNA nucleotide sequences respectively complementary and homologous to the target sequences.

Fire *et al.*, 1998 describe specific genetic interference by experimental introduction of double-stranded RNA in *Caenorhabditis elegans*. The importance of these findings for

- 25 functional genomics has been discussed (Wagner and Sun, 1998).

WO 99/32619 provides a process of introducing an RNA into a living cell to inhibit gene expression of a target gene in that cell. The process may be practiced *ex vivo* or *in vivo*. The RNA has a region with double-stranded structure. Inhibition is sequence-

- 30 specific in that the nucleotide sequences of the duplex region of the RNA and or a portion of the target gene are identical.

Waterhouse *et al.* 1998 describe that virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and anti-sense RNA. The sense and antisense RNA may be located in one transcript that has self-complementarity.

- 5 Hamilton *et al.* 1998 describes that a transgene with repeated DNA, i.e. inverted copies of its 5' untranslated region, causes high frequency, post-transcriptional suppression of ACC-oxidase expression in tomato.

WO 98/53083 describes constructs and methods for enhancing the inhibition of a
10 target gene within an organism which involve inserting into the gene silencing vector an inverted repeat sequence of all or part of a polynucleotide region within the vector.

WO 99/53050 provides methods and means for reducing the phenotypic expression of a nucleic acid of interest in eukaryotic cells, particularly in plant cells, by introducing
15 chimeric genes encoding sense and antisense RNA molecules directed towards the target nucleic acid, which are capable of forming a double stranded RNA region by base-pairing between the regions with the sense and antisense nucleotide sequence or by introducing the RNA molecules themselves. Preferably, the RNA molecules comprise simultaneously both sense and antisense nucleotide sequences.

20

WO 99/49029 relates generally to a method of modifying gene expression and to synthetic genes for modifying endogenous gene expression in a cell, tissue or organ of a transgenic organism, in particular to a transgenic animal of plant. Synthetic genes and genetic constructs, capable of forming a dsRNA which are capable of repressing,
25 delaying or otherwise reducing the expression of an endogenous gene or a target gene in an organism when introduced thereto are also provided.

WO 99/61631 relates to methods to alter the expression of a target gene in a plant using sense and antisense RNA fragments of the gene. The sense and antisense RNA
30 fragments are capable of pairing and forming a double-stranded RNA molecule, thereby altering the expression of the gene. The present invention also relates to plants, their progeny and seeds thereof obtained using these methods.

WO 00/01846 provides a method of identifying DNA responsible for conferring a particular phenotype in a cell which method comprises a) constructing a cDNA or genomic library of the DNA of the cell in a suitable vector in an orientation relative to (a) promoter(s) capable of initiating transcription of the cDNA or DNA to double stranded (ds) RNA upon binding of an appropriate transcription factor to the promoter(s); b) introducing the library into one or more of cells comprising the transcription factor, and c) identifying and isolating a particular phenotype of a cell comprising the library and identifying the DNA or cDNA fragment from the library responsible for conferring the phenotype. Using this technique, it is also possible to assign function to a known DNA sequence by a) identifying homologues of the DNA sequence in a cell, b) isolating the relevant DNA homologus(s) or a fragment thereof from the cell, c) cloning the homologue or fragment thereof into an appropriate vector in an orientation relative to a suitable promoter capable of initiating transcription of dsRNA from said DNA homologue or fragment upon binding of an appropriate transcription factor to the promoter and d) introducing the vector into the cell from step a) comprising the transcription factor.

WO 00/44914 also describes composition and methods for in vivo and in vitro attenuation of gene expression using double stranded RNA, particularly in zebrafish.

WO 00/49035 discloses a method for silencing the expression of an endogenous gene in a cell, the method involving overexpressing in the cell a nucleic acid molecule of the endogenous gene and an antisense molecule including a nucleic acid molecule complementary to the nucleic acid molecule of the endogenous gene, wherein the overexpression of the nucleic acid molecule of the endogenous gene and the antisense molecule in the cell silences the expression of the endogenous gene.

Smith et al., 2000 as well as WO 99/53050 described that intron containing dsRNA further increased the efficiency of silencing.

However, the prior art has not solved the problems associated with the efficient conversion of any nucleotide sequence of interest into a chimeric construct capable of

producing a dsRNA in eukaryotic cells, particularly in plant cells, and preferably in a way amenable to the processing of large number of nucleotide sequences.

These and other problems have been solved as described hereinafter in the different
5 embodiments and claims.

Summary of the invention.

It is an object of the invention to provide vectors comprising the following operably
10 linked DNA fragments a) an origin of replication allowing replication in microorganisms (1), preferably bacteria; particularly *Escherichia coli*; b) a selectable marker region (2) capable of being expressed in microorganisms, preferably bacteria; and c) a chimeric DNA construct comprising in sequence (i) a promoter or promoter region (3) capable of being recognized by RNA polymerases of a eukaryotic cell,
15 preferably a plant-expressible promoter; (ii) a first recombination site (4), a second recombination site (5), a third recombination site (6) and a fourth recombination site (7); and (iii) a 3' transcription terminating and polyadenylation region (8) functional in the eukaryotic cell; wherein the first recombination site (4) and the fourth recombination site (7) are capable of reacting with a same recombination site,
20 preferably are identical, and the second recombination site (5) and the third recombination site (6), are capable of reacting with a same recombination site, preferably are identical; and wherein the first recombination site (4) and the second recombination site (5) do not recombine with each other or with a same recombination site or the third recombination site (6) and the fourth recombination
25 site (7) do not recombine with each other or with a same recombination site.

Optionally the vector may further include additional elements such as: a second selectable marker gene (9) between the first (4) and second recombination site (5) and/or a third selectable marker gene (10) between the third (6) and fourth recombination site (7) and/or a region flanked by intron processing signals (11),
30 preferably an intron, functional in the eukaryotic cell, located between the second recombination site (5) and the third recombination site (6) and/or a fourth selectable marker gene (19), located between the second (5) and third recombination site (6) and/or left and right border T-DNA sequences flanking the chimeric DNA construct

plant, cells, preferably located between the left and the right T-DNA border sequences and/or an origin of replication capable of functioning in *Agrobacterium* spp. Selectable marker genes may be selected from the group consisting of an antibiotic resistance gene, a tRNA gene, an auxotrophic marker, a toxic gene, a phenotypic marker, an antisense oligonucleotide; a restriction endonuclease; a restriction endonuclease cleavage site, an enzyme cleavage site, a protein binding site, an a sequence complementary PCR primer. Preferably the first (4) and fourth recombination site (7) are *attR1* comprising the nucleotide sequence of SEQ ID No 4 and the second (5) and third (6) recombination site are *attR2* comprising the nucleotide sequence of SEQ ID No 5 or the first (4) and fourth recombination site (7) are *attP1* comprising the nucleotide sequence of SEQ ID No 10 and the second (5) and third (6) recombination site are *attP2* comprising the nucleotide sequence of SEQ ID No 11.

It is another objective of the invention to provide a kit comprising an acceptor vector according to invention, preferably further comprising at least one recombination protein capable of recombining a DNA segment comprising at least one of the recombination sites.

It is yet another objective of the invention to provide a method for making a chimeric DNA construct capable of expressing a dsRNA in a eukaryotic cell comprising the steps of

- a) combining in vitro:
 - i) an acceptor vector as herein before described;
 - ii) an insert DNA, preferably a linear or circular insert DNA, comprising a DNA segment of interest (12) flanked by
 - (a) a fifth recombination site (13) which is capable of recombining with the first (4) or fourth recombination site (7) on the vector; and
 - (b) a sixth recombination site (14) which is capable of recombining with the second (5) or third recombination site (6) on the vector;
 - iii) at least one site specific recombination protein capable of recombining the first (4) or fourth (7) and the fifth recombination site (13) and the second (5) or third (6) and the sixth recombination site (14);

- b) allowing recombination to occur in the presence of at least one recombination protein, preferably selected from Int and IHF and (ii) Int, Xis, and IHF, so as to produce a reaction mixture comprising product DNA molecules, the product DNA molecule comprising in sequence:
- 5 i) the promoter or promoter region (3) capable of being recognized by RNA polymerases of the eukaryotic cell;
 - ii) a recombination site (15) which is the recombination product of the first (4) and the fifth recombination site (13);
 - iii) the DNA fragment of interest (12);
 - 10 iv) a recombination site (16) which is the recombination product of the second (4) and the sixth recombination site (14);
 - v) a recombination site (17) which is the recombination product of the third (5) and the sixth recombination site (14);
 - vi) the DNA fragment of interest in opposite orientation (12);
 - 15 vii) a recombination site (18) which is the recombination product of the fourth (7) and the fifth recombination site (13); and
 - viii) the 3' transcription terminating and polyadenylation region (8) functional in the eukaryotic cell;
- c) selecting the product DNA molecules, preferably in vivo.

20

The method allows that multiple insert DNAs comprising different DNA fragments of interest are processed simultaneously.

- The invention also provides a method for preparing a eukaryotic non-human
- 25 organism, preferably a plant, wherein the expression of a target nucleic acid of interest is reduced or inhibited, the method comprising:

- a) preparing a chimeric DNA construct capable of expressing a dsRNA in cells of the eukaryotic non-human organism according to methods of the invention;
- 30 b) introducing the chimeric DNA construct in cells of the eukaryotic non-human organism; and
- c) isolating the transgenic eukaryotic organism

It is also an objective of the invention to provide a method for isolating a nucleic acid molecule involved in determining a particular trait

- a) preparing a library of chimeric DNA constructs capable of expressing a dsRNA in cells of the eukaryotic non-human organism according to any one of the methods of the invention;
- b) introducing individual representatives of the library of chimeric DNA constructs in cells of the eukaryotic non-human organism;
- c) isolating a eukaryotic organism exhibiting the particular trait; and isolating the nucleic acid molecule.

The invention also provides a eukaryotic non-human organism, preferably a plant comprising a chimeric DNA construct obtainable through the methods of the invention.

15 Brief description of the figures.

Figure 1. Schematic representation of vectors and method used in a preferred embodiment of the invention.

Figure 1A: A nucleic acid of interest (12) is amplified by PCR using primers comprising two different recombination sites (13, 14) which cannot react with each other or with the same other recombination site. This results in "insert DNA" wherein the nucleic acid of interest (12) is flanked by two different recombination sites (13, 14).

Figure 1B. Using at least one recombination protein, the insert DNA is allowed to recombine with the acceptor vector between the recombination sites, whereby the first (4) and fourth recombination site (7) react with one of the recombination sites (13) flanking the PCR amplified DNA of interest (12) and the second (5) and third (6) recombination site on the acceptor vector recombine with the other recombination site (14) flanking the DNA of interest (12). The desired product DNA can be isolated by selecting for loss of the selectable marker genes (9) and (10) located between respectively the first (4) and second (5) recombination sites and the third (6) and fourth (7) recombination sites. Optionally, an additional selectable marker gene may

be included between the second (5) and third (6) recombination site to allow selection for the presence of this selectable marker gene and consequently for the optional intron sequence, which is flanked by functional intron processing signal sequences (11). The acceptor vector, as well as the product vector further comprises a origin of replication (Ori; (1)) and a selectable marker gene (2) to allow selection for the presence of the plasmid.

This result in a chimeric DNA construct with the desired configuration comprising a eukaryotic promoter region (3); a recombination site (15) produced by the recombination between recombination sites (4) and (13); a first copy of the DNA of interest (12); a recombination site (16) produced by the recombination between recombination sites (5) and (14); optionally an intron sequence flanked by intron processing signals (11); a recombination site (17) produced by the recombination between recombination sites (6) and (14); a second copy of the DNA of interest (12) in opposite orientation to the first copy of the DNA of interest; a recombination site (18) produced by the recombination between recombination sites (7) and (13); a eukaryotic transcription terminator and polyadenylation signal (8).

Figure 2A: A nucleic acid of interest (12) is amplified by PCR using primers comprising two different recombination sites which upon recombination with the recombination sites on an intermediate vector (Figure 2B) will yield recombination sites compatible with the first (4) and fourth (5) and with the second (6) and third (7) recombination site on the acceptor vector respectively.

Figure 2B: The insert DNA obtained in Figure 2A is allowed to recombine with the intermediate vector in the presence of at least one recombination protein to obtain an intermediate DNA wherein the DNA of interest (12) is flanked by two different recombination sites (13, 14) and which further comprises an origin of replication (1) and a selectable marker gene (2).

Figure 2C: The intermediate DNA is then allowed to recombine with the acceptor vector using at least one second recombination protein (basically as described for Figure 1B).

Figure 3: Schematic representation of the acceptor vector "pHELLSGATE"

Figure 4: Schematic representation of the acceptor vectors "pHELLSGATE 8"

5 "pHELLSGATE 11" and "pHELLSGATE 12".

Detailed description of preferred embodiments.

The current invention is based on the unexpected finding by the inventors that recombinational cloning was an efficient one-step method to convert a nucleic acid
10 fragment of interest into a chimeric DNA construct capable of producing a dsRNA transcript comprising a sense and antisense nucleotide sequence capable of being expressed in eukaryotic cells. The dsRNA molecules are efficient effectors of gene-silencing. These methods improves the efficiency problems previously encountered to produce chimeric DNAs with long inverted repeats.

15

Thus, in a first embodiment, the invention provides a method for making a chimeric DNA construct or chimeric gene capable of expressing an RNA transcript in a eukaryotic cell, the RNA being capable of internal basepairing between a stretch of nucleotides corresponding to a nucleic acid of interest and its complement (i.e. the
20 stretch of nucleotides in inverted orientation) located elsewhere in the transcript (and thus forming a hairpin RNA) comprising the following steps:

1. Providing an "acceptor vector" comprising the following operably linked DNA fragments:
 - a) an origin of replication allowing replication in a host cell (1),
 - 25 b) a selectable marker region (2) capable of being expressed in the host cell; and
 - c) a chimeric DNA construct comprising in sequence:
 - i) a promoter or promoter region (3) capable of being recognized by RNA polymerases of a eukaryotic cell;
 - ii) a first recombination site (4), a second recombination site (5), a third
30 recombination site (6) and a fourth recombination site (7) whereby
(1) the first (4) and fourth recombination site (7) are capable of reacting with the same other recombination site and preferably are identical to each other;

- (2) the second (5) and third (6) recombination site are also capable of reacting with the same other recombination site and preferably are identical to each other
- (3) the first (4) and second (5) recombination site do not recombine with each other or with the same other recombination site; and
- (4) the third (6) and fourth (7) recombination site do not recombine with each other or with the same other recombination site; and
- iii) a 3' transcription terminating and polyadenylation region (8) functional in a eukaryotic cell.
2. Providing an "insert DNA" comprising the DNA segment of interest (12) flanked by
- a) a fifth recombination site (13) which is capable of recombining with the first (4) or fourth (7) recombination site but preferably not with the second (5) or third (6) recombination site;
- b) a sixth recombination site (14) which is capable of recombining with the second (5) or third (6) recombination site but preferably not with the first (4) or fourth (7) recombination site.
3. Combining *in vitro* the insert DNA and the acceptor vector in the presence of at least one specific recombination protein and allowing the recombination to occur to produce a reaction mixture comprising inter alia "product DNA" molecules which comprise in sequence
- i) the promoter or promoter region (3) capable of being recognized by RNA polymerases of a eukaryotic cell;
- ii) a recombination site (15) which is the recombination product of the first (4) and fifth recombination site (13);
- iii) a first copy of the DNA fragment of interest (12);
- iv) a recombination site (16) which is the recombination product of the second (4) and the sixth recombination site (14);
- v) a recombination site (17) which is the recombination product of the third (5) and the sixth recombination site (14);
- vi) a second copy of the DNA fragment of interest in opposite orientation (12) with regard to the first copy ;

- vii) a recombination site (18) which is the recombination product of the fourth (7) and the fifth recombination site (13); and
- viii) a 3' transcription terminating and polyadenylation region (8) functional in a eukaryotic cell;

5

4. Selecting the product DNA molecules.

This method is schematically outlined in Figure 1, with non-limiting examples of recombination sites and selectable markers.

10

As used herein, a "host cell" is any prokaryotic or eukaryotic organism that can be a recipient for the acceptor vector or the product DNA. Conveniently, the host cell will be a *Escherichia coli* strain commonly used in recombinant DNA methods.

- 15 A "recombination protein" is used herein to collectively refer to site specific recombinases and associated proteins and/or co-factors. Site specific recombinases are enzymes that are present in some viruses and bacteria and have been characterized to have both endonuclease and ligase properties. These recombinases (along with associated proteins in some cases) recognize specific sequences of bases in DNA and
- 20 exchange the DNA segments flanking those segments. Various recombination proteins are described in the art(see WO 96/40724 herein incorporated by reference in its entirety, at least on page 22 to 26).

Examples of such recombinases include Cre from bacteriophage P1 and Integrase from

25 bacteriophage lambda.

Cre is a protein from bacteriophage P1 (Abremski and Hoess, 1984) which catalyzes the exchange between 34 bp DNA sequences called *loxP* sites (see Hoess et al., 1986. Cre is available commercially (Novagen, Catalog 69247-1).

30

Integrase (Int) is a protein from bacteriophage lambda which mediates the integration of the lambda genome into the *E. coli* chromosome. The bacteriophage lambda Int recombinational proteins promote irreversible recombination between its substrate *att*

sites as part of the formation or induction of a lysogenic state. Reversibility of the recombination reactions results from two independent pathways for integrative or excisive recombination. Cooperative and competitive interactions involving four proteins (Int, Xis, IHF and FIS) determine the direction of recombination. Integrative
 5 recombination involves the Int and IHF proteins and *attP* (240bp) and *attB* (25b) recombination sites. Recombination results in the formation of two new sites: *attL* and *attR*. A commercial preparation comprising Int and IHF proteins is commercially available (BP clonase[™]; Life Technologies). Excisive recombination requires Int, IHF, and Xis and sites *attL* and *attR* to generate *attP* and *attB*. A commercial preparation
 10 comprising Int, IHF and Xis proteins is commercially available (LR clonase[™]; Life Technologies).

A "recombination site" as used herein refers to particular DNA sequences, which a recombinase and possibly associated proteins recognizes and binds. The
 15 recombination site recognized by Cre recombinase is *loxP* which is a 34 base pair sequence comprised of two 13 base pair inverted repeats (serving as recombinase binding sites) flanking an 8 base pair core sequence. The recombination sites *attB*, *attP*, *attL* and *attR* are recognized by lambda integrase. *AttB* is an approximately 25 base pair sequence containing two 9 base pair core-type Int binding sites and a 7 base
 20 pair overlap region. *AttP* is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins IHF, FIS and Xis (Landy 1993). Each of the *att* sites contains a 15 bp core sequence with individual sequence elements of functional significance lying within, outside and across the boundaries of this common core (Landy, 1989) Efficient
 25 recombination between the various *att* sites requires that the sequence of the central common region is substantially identical between the recombining partners. The exact sequence however is modifiable as disclosed in WO 96/40724 and the variant recombination sites selected from

- i) *attB1*: AGCCTGCTTTTTTGTACAAACTTGT (SEQ ID No 1);
- 30 ii) *attB2*: AGCCTGCTTCTTGTACAAACTTGT (SEQ ID No 2);
- iii) *attB3*: ACCCAGCTTCTTGTACAAACTTGT (SEQ ID No 3);
- iv) *attR1*: GTTCAGCTTTTTTGTACAAACTTGT (SEQ ID No 4);
- v) *attR2*: GTTCAGCTTCTTGTACAAACTTGT (SEQ ID No 5);

- vi) *attR3*: GTTCAGCTTTCTTGTACAAAGTTGG (SEQ ID No 6);
- vii) *attL1*: AGCCTGCTTTTTTGTACAAAGTTGG (SEQ ID No 7);
- viii) *attL2*: AGCCTGCTTTCTTGTACAAAGTTGG (SEQ ID No 8);
- ix) *attL3*: ACCCAGCTTTCTTGTACAAAGTTGG (SEQ ID No 9);
- 5 x) *attP1*: GTTCAGCTTTTTTGTACAAAGTTGG (SEQ ID No 10); or
- xi) *attP2,P3*: GTTCAGCTTTCTTGTACAAAGTTGG (SEQ ID No 11)

allow more flexibility in the choice of suitable pairs or recombination sites which are capable to recombine (as indicated by their index number).

- 10 It will be clear to the skilled artisan that a correspondence is required between the recombination site(s) used and the recombination proteins used.

In one embodiment the following combinations of recombination sites for the acceptor vector are present in the acceptor vector:

- 15 - the first (4) and fourth (7) recombination sites are identical and comprise *attP1* comprising the nucleotide sequence of SEQ ID No 10 and the second (5) and third (6) recombination site are also identical and comprise *attP2* comprising the nucleotide sequence of SEQ ID No 11; or
- the first (4) and fourth (7) recombination sites are identical and comprise *attR1*
- 20 comprising the nucleotide sequence of SEQ ID No 4 and the second (5) and third (6) recombination site are also identical and comprise *attR2* comprising the nucleotide sequence of SEQ ID No 5; and

the following combinations of recombination sites for the insert DNA are used:

- the fifth (13) recombination site comprises *attB1* comprising the nucleotide
- 25 sequence of SEQ ID No 1 and the sixth (14) recombination site comprises *attB2* comprising the nucleotide sequence of SEQ ID No 2, the combination being suitable for recombination with the first acceptor vector mentioned above; or
- the fifth (13) recombination site comprises *attL1* comprising the nucleotide sequence of SEQ ID No 7 and the sixth (14) recombination site comprises *attL2*
- 30 comprising the nucleotide sequence of SEQ ID No 8, the combination being suitable for recombination with the second acceptor vector mentioned above.

It has been unexpectedly found that product DNA molecules (resulting from recombination between the above mentioned second acceptor vector with attR recombination sites (such as pHELLSGATE 8) and insert DNA flanked by attL recombination sites) wherein the gene inserts in both orientations are flanked by attB recombination sites are more effective in silencing of the target gene(both quantitatively and qualitatively) than product DNA molecules (resulting from recombination between the above mentioned first acceptor vector with attP recombination sites (such as pHELLSGATE or pHELLSGATE 4) and insert DNA flanked by attB recombination sites) wherein the gene inserts in both orientations are flanked by attL recombination sites. Although not intending to limit the invention to a particular mode of action it is thought that the greater length of the attL sites and potential secondary structures therein may act to inhibit transcription yielding the required dsRNA to a certain extent. However, acceptor vectors such as the above mentioned first acceptor vectors with attP sites may be used when target gene silencing to a lesser extent would be useful or required.

The dsRNA obtained by the chimeric DNA construct made according to the invention may be used, to silence a nucleic acid of interest, i.e. reduce its phenotypic expression, in a eukaryotic organism, particularly a plant, either directly or by transcription of the chimeric DNA construct in the cells of the eukaryotic organism. When this is the case, the following considerations may apply.

The length of the nucleic acid of interest (12) may vary from about 10 nucleotides (nt) up to a length equaling the length (in nucleotides) of the target nucleic acid whose phenotypic expression is to be reduced. Preferably the total length of the sense nucleotide sequence is at least 10 nt, or at least 19 nt or at least 21 nt or at least 25 nt, or at least about 50 nt, or at least about 100 nt, or at least about 150 nt, or at least about 200 nt, or at least about 500 nt. It is expected that there is no upper limit to the total length of the sense nucleotide sequence, other than the total length of the target nucleic acid. However for practical reason (such as e.g. stability of the chimeric genes) it is expected that the length of the sense nucleotide sequence should not exceed 5000 nt, particularly should not exceed 2500 nt and could be limited to about 1000 nt.

It will be appreciated that the longer the total length of the nucleic acid of interest (12), the less stringent the requirements for sequence identity between the nucleic acid of interest and the corresponding sequence in the target gene. Preferably, the nucleic acid of interest should have a sequence identity of at least about 75% with the
5 corresponding target sequence, particularly at least about 80 %, more particularly at least about 85%, quite particularly about 90%, especially about 95%, more especially about 100%, quite especially be identical to the corresponding part of the target nucleic acid. However, it is preferred that the nucleic acid of interest always includes a sequence of about 10 consecutive nucleotides, particularly about 25 nt, more
10 particularly about 50 nt, especially about 100 nt, quite especially about 150 nt with 100% sequence identity to the corresponding part of the target nucleic acid. Preferably, for calculating the sequence identity and designing the corresponding sense sequence, the number of gaps should be minimized, particularly for the shorter sense sequences.

15

For the purpose of this invention, the "sequence identity" of two related nucleotide or amino acid sequences, expressed as a percentage, refers to the number of positions in the two optimally aligned sequences which have identical residues (x100) divided by the number of positions compared. A gap, i.e. a position in an alignment where a
20 residue is present in one sequence but not in the other is regarded as a position with non-identical residues. The alignment of the two sequences is performed by the Needleman and Wunsch algorithm (Needleman and Wunsch 1970) The computer-assisted sequence alignment above, can be conveniently performed using standard software program such as GAP which is part of the Wisconsin Package Version 10.1
25 (Genetics Computer Group, Madison, Wisconsin, USA) using the default scoring matrix with a gap creation penalty of 50 and a gap extension penalty of 3. Sequences are indicated as "essentially similar" when such sequence have a sequence identity of at least about 75%, particularly at least about 80 %, more particularly at least about 85%, quite particularly about 90%, especially about 95%, more especially about
30 100%, quite especially are identical. It is clear than when RNA sequences are the to be essentially similar or have a certain degree of sequence identity with DNA sequences, thymine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence.

The "insert DNA" may conveniently be provided using DNA amplification procedures, such as PCR, of the nucleic acid of interest, using as primers oligonucleotide sequences incorporating appropriate recombination sites as well as oligonucleotide sequences appropriate for the amplification of the nucleic acid of interest. However, alternative methods are available in the art to provide the nucleic acid of interest with the flanking recombination sites, including but not limited to covalently linking oligonucleotides or nucleic acid fragments comprising such recombination sites to the nucleic acid(s) of interest using ligase(s).

10

The providing of the appropriate flanking recombination sites to the nucleic acid may also proceed in several steps. E.g. in a first step the flanking sites provided to the nucleic acid of interest may be such that upon recombination with the recombination sites in an intermediate vector new recombination sites are created flanking the nucleic acid of interest, now compatible for recombination with the acceptor vector. This scheme is outlined in Figure 2, with non-limiting examples of recombination sites and selectable markers. It goes without saying that the insert DNA may be in a circular form or in a linear form.

As used herein, an "origin of replication" is a DNA fragment which allows replication of the acceptor vector in microorganisms, preferably bacteria, particularly *E. coli* strains, and ensures that upon multiplication of the microorganism, the daughter cells receive copies of the acceptor vector.

"Selectable marker (gene)" is used herein to indicate a DNA segment which allows to select or screen for the presence or absence of that DNA segment under suitable conditions. Selectable markers include but are not limited to

- (1) DNA segments that encode products which provide resistance against otherwise toxic compounds (e.g. antibiotic resistance genes, herbicide resistance genes)
- (2) DNA segments encoding products which are otherwise lacking in the recipient cell (e.g. tRNA genes, auxotrophic markers)

30

- (3) DNA segments encoding products which suppress the activity of a gene product;
- (4) DNA segments encoding products which can readily be identified (e.g. β -galactosidase, green fluorescent protein (GFP), β -glucuronidase (GUS));
- 5 (5) DNA segments that bind products which are otherwise detrimental to cell survival and/or function;
- (6) DNA segments that are capable of inhibiting the activity of any of the DNA segments described in Nos 1 to 5 (e.g. antisense oligonucleotides);
- (7) DNA segments that bind products that modify a substrate (e.g. restriction
10 endonuclease);
- (8) DNA segments that can be used to isolate a desired molecule (e.g. specific protein binding sites);
- (9) DNA segments that encode a specific nucleotide sequence which can be otherwise non-functional (e.g. for PCR amplification of subpopulations
15 of molecules);
- (10) DNA segments, which when absent, directly or indirectly confer sensitivity to particular compound(s);
- (11) DNA segments, which when absent, directly or indirectly confer resistance to particular compound(s);

20

Preferred first selectable markers (2) are antibiotic resistance genes. A large number of antibiotic resistance genes, particularly which can be used in bacteria, are available in the art and include but are not limited to aminoglycoside phosphotransferase I and II, chloramphenicol acetyltransferase, beta-lactamase, aminoglycoside
25 adenosyltransferase.

Preferred second selectable marker (9) and third selectable markers (10) are selectable markers allowing a positive selection when absent or deleted after recombination (i.e. in the product DNA) such as but not limited to *ccdB* gene the product of which
30 interferes with *E. coli* DNA gyrase and thereby inhibits growth of most *E. coli* strains. Preferably, the second and third marker are identical.

In one embodiment of the invention, the acceptor comprises a fourth selectable marker (19) between the second (5) and third (6) recombination site, preferably a marker allowing positive selection for the presence thereof, such as a antibiotic resistance gene, e.g. chloramphenicol resistance gene. Preferably, the fourth selectable
5 marker should be different from first selectable marker and different from the second and third selectable marker. The presence of a fourth selectable marker allows to select or screen for the retention of the DNA region between the second (5) and third (6) recombination site in the product DNA, thereby increasing the efficiency with which the desired product DNAs having the nucleic acid of interest cloned in inverted
10 repeat and operably linked to eukaryotic expression signals may be obtained.

However, it has been found that with most of the acceptor vectors tested, the presence of a selectable marker is not required and has little influence on the ratio of expected and desired product DNA molecules (which usually exceeds about 90% of obtained product DNA molecules) to undesired product DNA molecules.

15 It goes without saying that a person skilled in the art has a number of techniques available for recognizing the expected and desired product DNA molecules, such as but not limited to restriction enzyme digests or even determining the nucleotide sequence of the recombination product.

20 In another embodiment of the invention, the acceptor vector further comprises a pair of intron processing signals (11) or an intron sequence functional in the eukaryotic cell, preferably located between the second (5) and third (6) recombination site. However, the pair of intron processing signals or the intron may also be located
25 elsewhere in the chimeric construct between the promoter or promoter region (3) and the terminator region (8). As indicated in the background art, this will improve the efficiency with which the chimeric DNA construct encoding the dsRNA will be capable of reducing the phenotypic expression of the target gene in the eukaryotic cell. A particularly preferred intron functional in cells of plants is the *pdk* intron
30 (*Flaveria trinervia* pyruvate orthophosphate dikinase intron 2 ; see WO99/53050 incorporated by reference). The fourth selectable marker (19) may be located between the intron processing signals or within the intron (if these are located between the

second and third recombination site), but may also be located adjacent to the intron processing signals or the intron.

A person skilled in the art will recognize that the product DNA molecules resulting from a recombination with an acceptor vector as herein described which comprise a region between the second (5) and third (6) recombination will fall into two classes which can be recognized by virtue of the orientation of that intervening region. In the embodiments wherein the acceptor vector also comprises an intron, the different orientation may necessitate an additional step of identifying the correct orientation. To avoid this additional step, the acceptor vector may comprise an intron which can be spliced out independent of its orientation (such as present in pHELLSGATE 11) or the acceptor vector may comprise an spliceable intron in both orientations (such as present in pHELLSGATE 12).

As used herein, the term "promoter" denotes any DNA which is recognized and bound (directly or indirectly) by a DNA-dependent RNA-polymerase during initiation of transcription. A promoter includes the transcription initiation site, and binding sites for transcription initiation factors and RNA polymerase, and can comprise various other sites (e.g., enhancers), at which gene expression regulatory proteins may bind.

20

The term "regulatory region", as used herein, means any DNA, that is involved in driving transcription and controlling (i.e., regulating) the timing and level of transcription of a given DNA sequence, such as a DNA coding for a protein or polypeptide. For example, a 5' regulatory region (or "promoter region") is a DNA sequence located upstream (i.e., 5') of a coding sequence and which comprises the promoter and the 5'-untranslated leader sequence. A 3' regulatory region is a DNA sequence located downstream (i.e., 3') of the coding sequence and which comprises suitable transcription termination (and/or regulation) signals, including one or more polyadenylation signals.

30

As used herein, the term "plant-expressible promoter" means a DNA sequence which is capable of controlling (initiating) transcription in a plant cell. This includes any promoter of plant origin, but also any promoter of non-plant origin which is capable

of directing transcription in a plant cell, i.e., certain promoters of viral or bacterial origin such as the CaMV35S, the subterranean clover virus promoter No 4 or No 7, or T-DNA gene promoters but also tissue-specific or organ-specific promoters including but not limited to seed-specific promoters (e.g., WO89/03887), organ-primordia
5 specific promoters (An et al., 1996), stem-specific promoters (Keller et al., 1988), leaf specific promoters (Hudspeth et al., 1989), mesophyl-specific promoters (such as the light-inducible Rubisco promoters), root-specific promoters (Keller et al., 1989), tuber-specific promoters (Keil et al., 1989), vascular tissue specific promoters (Peleman et al., 1989), stamen-selective promoters (WO 89/10396, WO 92/13956), dehiscence
10 zone specific promoters (WO 97/13865) and the like.

The acceptor vector may further comprise a selectable marker for expression in a eukaryotic cell. Selectable marker genes for expression in eukaryotic cells are well known in the art, including but not limited to chimeric marker genes. The chimeric
15 marker gene can comprise a marker DNA that is operably linked at its 5' end to a promoter, functioning in the host cell of interest, particularly a plant-expressible promoter, preferably a constitutive promoter, such as the CaMV 35S promoter, or a light inducible promoter such as the promoter of the gene encoding the small subunit of Rubisco; and operably linked at its 3' end to suitable plant transcription 3' end
20 formation and polyadenylation signals. It is expected that the choice of the marker DNA is not critical, and any suitable marker DNA can be used. For example, a marker DNA can encode a protein that provides a distinguishable colour to the transformed plant cell, such as the A1 gene (Meyer et al., 1987), can provide herbicide resistance to the transformed plant cell, such as the *bar* gene, encoding resistance to
25 phosphinothricin (EP 0,242,246), or can provide antibiotic resistance to the transformed cells, such as the *aac(6')* gene, encoding resistance to gentamycin (WO94/01560).

The acceptor vector may also further comprise left and right T-DNA border sequences
30 flanking the chimeric DNA construct, and may comprise an origin of replication functional in *Agrobacterium* spp. and/or a DNA region of homology with a helper Ti-plasmid as described in EP 0 116 718.

The efficiency and ease by which any nucleic acid of interest may be converted into a chimeric DNA construct comprising two copies of the nucleic acid of interest in inverted repeat and operably linked to eukaryotic 5' and 3' regulatory regions using the means and methods according to the invention, makes these particularly apt for
5 automation and high throughput analysis.

It will be clear to the person skilled in the art that the acceptor vectors as hereinbefore described can be readily adapted to provide a vector which can be used to produce *in vitro* large amounts of double stranded RNA or RNAi comprising a complementary
10 sense and antisense portion essentially similar to a target gene of choice as described elsewhere in this application, by exchanging the promoter capable of being expressed in a eukaryotic cell for a promoter recognized by any RNA polymerase. Very suitable promoters to this end are the promoters recognized by bacteriophage single subunit RNA polymerases such as the promoters recognized by bacteriophage single subunit
15 RNA polymerase such as the RNA polymerases derived from the E. coli phages T7, T3, ϕ I, ϕ II, W31, H, Y, A1, 122, cro, C21, C22, and C2; Pseudomonas putida phage gh-1; Salmonella typhimurium phage SP6; Serratia marcescens phage IV; Citrobacter phage ViIII; and Klebsiella phage No.11 [Hausmann, Current Topics in Microbiology and Immunology, 75: 77-109 (1976); Korsten et al., J. Gen Virol. 43: 57-73 (1975);
20 Dunn et al., Nature New Biology, 230: 94-96 (1971); Towle et al., J. Biol. Chem. 250: 1723-1733 (1975); Butler and Chamberlin, J. Biol. Chem., 257: 5772-5778 (1982)]. Examples of such promoters are a T3 RNA polymerase specific promoter and a T7 RNA polymerase specific promoter, respectively. A T3 promoter to be used as a first promoter in the CIG can be any promoter of the T3 genes as described by McGraw et
25 al, Nucl. Acid Res. 13: 6753-6766 (1985). Alternatively, a T3 promoter may be a T7 promoter which is modified at nucleotide positions -10, -11 and -12 in order to be recognized by T3 RNA polymerase [(Klement et al., J. Mol. Biol. 215, 21-29(1990)]. A preferred T3 promoter is the promoter having the "consensus" sequence for a T3 promoter, as described in US Patent 5,037,745. A T7 promoter which may be used
30 according to the invention, in combination with T7 RNA polymerase, comprises a promoter of one of the T7 genes as described by Dunn and Studier, J. Mol. Biol. 166: 477-535 (1983). A preferred T7 promoter is the promoter having the "consensus"

sequence for a T7 promoter, as described by Dunn and Studier (supra). Thus, the invention also provides an acceptor vector comprising

- a) origin of replication allowing replication in a host cell (1),
 - b) a selectable marker region (2) capable of being expressed in the host cell; and
 - 5 c) a chimeric DNA construct comprising in sequence:
 - i) a promoter or promoter region (3) capable of being recognized by a bacteriophage single subunit RNA polymerase;
 - ii) a first recombination site (4), a second recombination site (5), a third recombination site (6) and a fourth recombination site (7) whereby
- 10 (1) the first (4) and fourth recombination site (7) are capable of reacting with the same other recombination site and preferably are identical to each other;
- (2) the second (5) and third (6) recombination site are also capable of reacting with the same other recombination site and preferably are
- 15 identical to each other
- (3) the first (4) and second (5) recombination site do not recombine with each other or with the same other recombination site; and
- (4) the third (6) and fourth (7) recombination site do not recombine with each other or with the same other recombination site; and
- 20 (5) a 3' transcription terminating and polyadenylation region (8) functional in a eukaryotic cell.

The acceptor vector may be used to convert a DNA fragment of interest into an inverted repeat structure as described elsewhere in the application and dsRNA can be

25 produced in large amounts by contacting the acceptor vector DNA with the appropriate bacteriophage single subunit RNA polymerase under conditions well known to the skilled artisan. The so-produced dsRNA can then be used for delivery into cells prone to gene silencing, such as plant cells, fungal cells or animal cells. dsRNA may be introduced in animal cells via liposomes or other transfection agents

30 (e.g. Clonfection transfection reagent or the CalPhos Mammalian transfection kit from ClonTech) and could be used for methods of treatment of animals, including humans, by silencing the appropriate target genes.

The acceptor vectors may also be equipped with any prokaryotic promoter suitable for expression of dsRNA in a particular prokaryotic host. The prokaryotic host can be used as a source of dsRNA, e.g. by feeding it to an animal, such as a nematode, in which the silencing of the target gene is envisioned.

5

The promoter capable of expression in eukaryotic cell may also be a promoter capable of expression in a mammalian cell and vectors according to the invention may transiently be delivered using a retroviral delivery system or other animal transfection system.

10

In another embodiment of the invention, a method is provided for making a eukaryotic organism, particularly a plant, wherein the phenotypic expression of a target nucleic acid of interest is reduced or inhibited, comprising the steps of preparing a chimeric DNA construct comprising a nucleic acid of interest (12) comprising a nucleotide sequence of at least 19 bp or 25 bp having at least 70% sequence identity to the target nucleic acid of interest and capable of expressing a dsRNA in cells of the eukaryotic organism, particularly a plant according to the methods of the current invention and introducing the chimeric DNA construct in cells of the eukaryotic organism, and isolating eukaryotic organism transgenic for the chimeric DNA construct.

As used herein, "phenotypic expression of a target nucleic acid of interest" refers to any quantitative trait associated with the molecular expression of a nucleic acid in a host cell and may thus include the quantity of RNA molecules transcribed or replicated, the quantity of post-transcriptionally modified RNA molecules, the quantity of translated peptides or proteins, the activity of such peptides or proteins.

A "phenotypic trait" associated with the phenotypic expression of a nucleic acid of interest refers to any quantitative or qualitative trait, including the trait mentioned, as well as the direct or indirect effect mediated upon the cell, or the organism containing that cell, by the presence of the RNA molecules, peptide or protein, or posttranslationally modified peptide or protein. The mere presence of a nucleic acid in a host cell, is not considered a phenotypic expression or a phenotypic trait of that

nucleic acid, even though it can be quantitatively or qualitatively traced. Examples of direct or indirect effects mediated on cells or organisms are, e.g., agronomically or industrial useful traits, such as resistance to a pest or disease; higher or modified oil content etc.

5

As used herein, "reduction of phenotypic expression" refers to the comparison of the phenotypic expression of the target nucleic acid of interest to the eucaryotic cell in the presence of the RNA or chimeric genes of the invention, to the phenotypic expression of the target nucleic acid of interest in the absence of the RNA or chimeric genes of the invention. The phenotypic expression in the presence of the chimeric RNA of the invention should thus be lower than the phenotypic expression in absence thereof, preferably be only about 25%, particularly only about 10%, more particularly only about 5% of the phenotypic expression in absence of the chimeric RNA, especially the phenotypic expression should be completely inhibited for all practical purposes by the presence of the chimeric RNA or the chimeric gene encoding such an RNA.

A reduction of phenotypic expression of a nucleic acid where the phenotype is a qualitative trait means that in the presence of the chimeric RNA or gene of the invention, the phenotypic trait switches to a different discrete state when compared to a situation in which such RNA or gene is absent. A reduction of phenotypic expression of a nucleic acid may thus, *i.a.* be measured as a reduction in transcription of (part of) that nucleic acid, a reduction in translation of (part of) that nucleic acid or a reduction in the effect the presence of the transcribed RNA(s) or translated polypeptide(s) have on the eucaryotic cell or the organism, and will ultimately lead to altered phenotypic traits. It is clear that the reduction in phenotypic expression of a target nucleic acid of interest, may be accompanied by or correlated to an increase in a phenotypic trait.

30 As used herein a "target nucleic acid of interest" refers to any particular RNA molecule or DNA sequence which may be present in a eucaryotic cell, particularly a plant cell whether it is an endogenous nucleic acid, a transgenic nucleic acid, a viral nucleic acid, or the like.

Methods for making transgenic eukaryotic organisms, particularly plants are well known in the art. Gene transfer can be carried out with a vector that is a disarmed Ti-plasmid, comprising a chimeric gene of the invention, and carried by *Agrobacterium*.

5 This transformation can be carried out using the procedures described, for example, in EP 0 116 718. A particular kind of *Agrobacterium* mediated transformation methods are the so-called *in planta* methods, which are particularly suited for *Arabidopsis* spp. transformation (e.g. Clough and Bent 1998). Alternatively, any type of vector can be used to transform the plant cell, applying methods such as direct gene transfer (as

10 described, for example, in EP 0 233 247), pollen-mediated transformation (as described, for example, in EP 0 270 356, WO85/01856 and US 4,684,611), plant RNA virus-mediated transformation (as described, for example, in EP 0 067 553 and US 4,407,956), liposome-mediated transformation (as described, for example, in US 4,536,475), and the like. Other methods, such as microprojectile bombardment, as

15 described for corn by Fromm *et al.* (1990) and Gordon-Kamm *et al.* (1990), are suitable as well. Cells of monocotyledonous plants, such as the major cereals, can also be transformed using wounded and/or enzyme-degraded compact embryogenic tissue capable of forming compact embryogenic callus, or wounded and/or degraded immature embryos as described in WO92/09696. The resulting transformed plant cell

20 can then be used to regenerate a transformed plant in a conventional manner.

The obtained transformed plant can be used in a conventional breeding scheme to produce more transformed plants with the same characteristics or to introduce the chimeric gene for reduction of the phenotypic expression of a nucleic acid of interest

25 of the invention in other varieties of the same or related plant species, or in hybrid plants. Seeds obtained from the transformed plants contain the chimeric genes of the invention as a stable genomic insert.

In another embodiment the invention provides a method for isolating a nucleic acid

30 molecule involved in determining a particular phenotypic trait of interest. The method involves the following steps:

- a) preparing a library of chimeric DNA constructs capable of expressing a dsRNA in cells of the eukaryotic non-human organism using the methods and means described in the current invention;
- b) introducing individual representatives of this library of chimeric DNA
5 constructs in cells of the eukaryotic non-human organism, preferably by stable integration in their genome, particularly their nuclear genome;
- c) isolating a eukaryotic organism exhibiting the particular trait; and
- d) isolating the corresponding nucleic acid molecule present in the eukaryotic
10 organism with the trait of interest, preferably from the aforementioned library.

10 It goes without saying that the methods and means of the invention may be used to determine the function of an isolated nucleic acid fragment or sequence with unknown function, by converting a part or the whole of that nucleic acid fragment or sequence according to the methods of the invention into a chimeric construct capable
15 of making a dsRNA transcript when introduced in a eukaryotic cell, introducing that chimeric DNA construct into a eukaryotic organism to isolate preferably a number of transgenic organisms and observing changes in phenotypic traits.

The invention also provides acceptor vectors, as described in this specification as well
20 as kits comprising the such vectors.

It goes without saying that the vectors, methods and kits according to the invention may be used in all eukaryotic organisms which are prone to gene silencing including yeast, fungi, plants, animals such as nematodes, insects and arthropods, vertebrates
25 including mammals and humans.

Also provided by the invention are non-human organisms comprising chimeric DNA constructs comprising in sequence the following operably linked DNA fragments

- i) a promoter or promoter region (3) capable of being recognized by RNA
30 polymerases of the eukaryotic cell;
- ii) a recombination site (15) which is the recombination product of the first (4) recombination site on the acceptor vector and the fifth recombination site (13) flanking the DNA of interest;

- iii) a first DNA copy of the nucleic acid fragment of interest (12);
- iv) a recombination site (16) which is the recombination product of the second (4) recombination site on the acceptor vector and the sixth recombination site (14) flanking the DNA of interest;
- 5 v) a recombination site (17) which is the recombination product of the third (5) recombination site on the acceptor vector and the sixth recombination site (14) flanking the DNA of interest;
- vi) a second DNA copy of the nucleic acid fragment of interest in opposite orientation (12) compared to the first copy;
- 10 vii) a recombination site (18) which is the recombination product of the fourth (7) recombination site on the acceptor vector and the fifth recombination site (13) flanking the DNA of interest; and
- viii) a 3' transcription terminating and polyadenylation region (8) functional in a eukaryotic cell.

15

As used herein "comprising" is to be interpreted as specifying the presence of the stated features, integers, steps or components as referred to, but does not preclude the presence or addition of one or more features, integers, steps or components, or groups thereof. Thus, e.g., a nucleic acid or protein comprising a sequence of nucleotides or amino acids, may comprise more nucleotides or amino acids than the actually cited ones, i.e., be embedded in a larger nucleic acid or protein. A chimeric gene comprising a DNA region which is functionally or structurally defined, may comprise additional DNA regions etc.

- 25 The term "gene" means any DNA fragment comprising a DNA region (the "transcribed DNA region") that is transcribed into a RNA molecule (e.g., a mRNA) in a cell operably linked to suitable regulatory regions, e.g., a plant-expressible promoter. A gene may thus comprise several operably linked DNA fragments such as a promoter, a 5' leader sequence, a coding region, and a 3' region comprising a polyadenylation
- 30 site. A plant gene endogenous to a particular plant species (endogenous plant gene) is a gene which is naturally found in that plant species or which can be introduced in that plant species by conventional breeding. A chimeric gene is any gene which is not normally found in a plant species or, alternatively, any gene in which the

promoter is not associated in nature with part or all of the transcribed DNA region or with at least one other regulatory region of the gene.

The term "expression of a gene" refers to the process wherein a DNA region which is operably linked to appropriate regulatory regions, particularly to a promoter, is transcribed into an RNA which is biologically active i.e., which is either capable of interaction with another nucleic acid or which is capable of being translated into a polypeptide or protein. A gene is the to encode an RNA when the end product of the expression of the gene is biologically active RNA, such as e.g. an antisense RNA, a ribozyme or a replicative intermediate. A gene is the to encode a protein when the end product of the expression of the gene is a protein or polypeptide.

A nucleic acid is "capable of being expressed", when the nucleic acid, when introduced in a suitable host cell, particularly in a plant cell, can be transcribed (or replicated) to yield an RNA, and/or translated to yield a polypeptide or protein in that host cell.

The following non-limiting Examples describe the construction of acceptor vectors and the application thereof for the conversion of nucleic acid fragments of interest into chimeric DNA constructs capable of expressing a dsRNA transcript in eukaryotic cells. Unless stated otherwise in the Examples, all recombinant DNA techniques are carried out according to standard protocols as described in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, NY and in Volumes 1 and 2 of Ausubel *et al.* (1994) *Current Protocols in Molecular Biology*, *Current Protocols*, USA. Standard materials and methods for plant molecular work are described in *Plant Molecular Biology Labfax* (1993) by R.D.D. Croy, jointly published by BIOS Scientific Publications Ltd (UK) and Blackwell Scientific Publications, UK. Other references for standard molecular biology techniques include Sambrook and Russell (2001) *Molecular Cloning: A Laboratory Manual*, Third Edition, Cold Spring Harbor Laboratory Press, NY, Volumes I and II of Brown (1998) *Molecular Biology LabFax*, Second Edition, Academic Press (UK). Standard materials and methods for polymerase chain reactions can be found in Dieffenbach and Dveksler (1995) *PCR Primer: A Laboratory Manual*, Cold Spring

Harbor Laboratory Press, and in McPherson et al. (2000) *PCR - Basics: From Background to Bench*, First Edition, Springer Verlag, Germany.

Throughout the description and Examples, reference is made to the following

5 sequences:

- | | | |
|----|---------------|---|
| | SEQ ID No 1: | core sequence of recombination site <i>attB1</i> |
| | SEQ ID No 2: | core sequence of recombination site <i>attB2</i> |
| | SEQ ID No 3: | core sequence of recombination site <i>attB3</i> |
| | SEQ ID No 4: | core sequence of recombination site <i>attR1</i> |
| 10 | SEQ ID No 5: | core sequence of recombination site <i>attR2</i> |
| | SEQ ID No 6: | core sequence of recombination site <i>attR3</i> |
| | SEQ ID No 7: | core sequence of recombination site <i>attL1</i> |
| | SEQ ID No 8: | core sequence of recombination site <i>attL2</i> |
| | SEQ ID No 9: | core sequence of recombination site <i>attL3</i> |
| 15 | SEQ ID No 10: | core sequence of recombination site <i>attP1</i> |
| | SEQ ID No 11: | core sequence of recombination sites <i>attP2,P3</i> |
| | SEQ ID No 12: | nucleotide sequence of chalcone synthase gene of <i>Arabidopsis</i> |
| | SEQ ID No 13: | nucleotide sequence of the acceptor vector "pHELLSGATE" |
| 20 | SEQ ID No 14: | oligonucleotide <i>attB1</i> "forward" primer used for amplification of 400bp and 200 bp CHS fragments. |
| | SEQ ID No 15: | oligonucleotide <i>attB2</i> "reverse" primer for amplification of the 400 bp CHS fragment. |
| | SEQ ID No 16: | oligonucleotide <i>attB2</i> "reverse" primer for amplification of the 200 bp CHS fragment. |
| 25 | SEQ ID No 17: | oligonucleotide <i>attB1</i> "forward" primer used for amplification of 100 bp CHS fragment. |
| | SEQ ID No 18: | oligonucleotide <i>attB2</i> "reverse" primer for amplification of the 100 bp CHS fragment. |
| 30 | SEQ ID No 19: | oligonucleotide <i>attB1</i> "forward" primer used for amplification of 50 bp CHS fragment. |
| | SEQ ID No 20: | oligonucleotide <i>attB2</i> "reverse" primer for amplification of the 50 bp CHS fragment. |

- SEQ ID No 21: oligonucleotide attB1 "forward" primer for amplification of the 25 bp CHS fragment.
- SEQ ID No 22: oligonucleotide attB2 "reverse" primer for the 25 bp fragment.
- 5 SEQ ID No 23: nucleotide sequence of the acceptor vector "pHELLSGATE 4"
- SEQ ID No 24: nucleotide sequence of the acceptor vector "pHELLSGATE 8"
- SEQ ID No 25: nucleotide sequence of the acceptor vector "pHELLSGATE 11"
- 10 SEQ ID No 26: nucleotide sequence of the acceptor vector "pHELLSGATE 12"

Examples

15

Example 1

Construction of the acceptor vector pHELLSGATE

With the completion of the *Arabidopsis* genome project, the advent of micro-array technology and the ever-increasing investigation into plant metabolic, perception, and response pathways, a rapid targeted way of silencing genes would be of major assistance. The high incidence and degree of silencing in plants transformed with chimeric genes containing simultaneously a sense and antisense nucleotide sequence, as well as a functional intron sequence suggested that such vectors could form the basis of a high-throughput silencing vector. However, one of the major obstacles in using such conventional cloning vectors for a large number of defined genes or a library of undefined genes would be cloning the hairpin arm sequences for each gene in the correct orientations.

25

Attempts to clone PCR products of sense and antisense arms together with the appropriately cut vector as a single step four-fragment ligation failed to give efficient or reproducible results. Therefore a construct (pHELLSGATE) was made to take advantage of Gateway™ (Life Technologies). With this technology, a PCR fragment is generated, bordered with recombination sites (*attB1* and *attB2*) which is directionally recombined, *in vitro*, into a plasmid containing two sets of suitable recombination

30

sites (*attP1* and *attP2* sites) using the commercially available recombination protein preparation.

The pHELLSGATE vector was designed such that a single PCR product from primers
5 with the appropriate *attB1* and *attB2* sites would be recombined into it simultaneously to form the two arms of the hairpin. The *ccdB* gene, which is lethal in standard *E.coli* strains such as DH5 α (but not in DB3.1), was placed in the locations to be replaced by the arm sequences, ensuring that only recombinants containing both arms would be recovered. Placing a chloramphenicol resistance gene within the
10 intron, gives a selection to ensure the retention of the intron in the recombinant plasmid.

pHELLSGATE comprises the following DNA fragments:

- a spectinomycin/streptomycin resistance gene (SEQ ID No 13 from the nucleotide
15 at position 7922 to the nucleotide sequence at 9985);
- a right T-DNA border sequence (SEQ ID No 13 from the nucleotide at position 10706 to the nucleotide sequence at 11324);
- a CaMV35S promoter (SEQ ID No 13 from the nucleotide at position 11674 to the nucleotide sequence at 13019);
- 20 • an *attP1* recombination site (complement of the nucleotide sequence of SEQ ID No 13 from the nucleotide at position 17659 to the nucleotide sequence at 17890);
- a *ccdB* selection marker (complement of the nucleotide sequence of SEQ ID No 13 from the nucleotide at position 16855 to the nucleotide at position 17610)
- an *attP2* recombination site (complement of the nucleotide sequence of SEQ ID No
25 13 from the nucleotide at position 16319 to the nucleotide at position 16551).
- *pdh* intron2 (SEQ ID No 13 from the nucleotide at position 14660 to the nucleotide at position 16258) flanked by the intron splice site (TACAG*TT (SEQ ID No 13 from the nucleotide at position 16254 to the nucleotide sequence at 16260) and the intron splice site (TG*GTAAG) (SEQ ID No 13 from the nucleotide
30 at position 14660 to the nucleotide sequence at 14667) and comprising a chloramphenicol resistance gene (SEQ ID No 13 from the nucleotide at position 15002 to the nucleotide at position 15661);

- an *attP2* recombination site (SEQ ID No 13 from the nucleotide at position 14387 to the nucleotide at position 14619)
- a *ccdB* selection marker (complement of the nucleotide sequence of SEQ ID No 13 from the nucleotide at position 13675 to the nucleotide at position 13980)
- 5 • an *attP1* recombination site (SEQ ID No 13 from the nucleotide at position 13048 to the nucleotide at position 13279)
- a octopine synthase gene terminator region (SEQ ID No 13 from the nucleotide at position 17922 to the nucleotide sequence at 18687);
- a chimeric marker selectable in plants comprising:
 - 10 • a nopaline synthase promoter (SEQ ID No 13 from the nucleotide at position 264 to the nucleotide sequence at 496);
 - a nptII coding region (SEQ ID No 13 from the nucleotide at position 497 to the nucleotide sequence at 1442); and
 - a nopaline synthase gene terminator (SEQ ID No 13 from the nucleotide at
 - 15 position 1443 to the nucleotide sequence at 2148);
 - a left T-DNA border sequence (SEQ ID No 13 from the nucleotide at position 2149 to the nucleotide sequence at 2706);
 - an origin of replication
 - a kanamycin resistance gene

20

The complete nucleotide sequence of pHELLSGATE is represented in the sequence listing (SEQ ID No 13) and a schematic figure can be found in Figure 3.

Example 2

- 25 **Use of the pHELLSGATE to convert nucleic acid fragments of interest into dsRNA producing chimeric silencing genes.**

To test the acceptor vector pHELLSGATE an about 400bp, 200bp, 100bp, 50 bp and 25 bp fragment of the *Arabidopsis thaliana* chalcone synthase isomerase coding sequence (Seq ID No 12) (having respectively the nucleotide sequence of SEQ ID No 12 from

30 the nucleotide at position 83 to the nucleotide at position 482; the nucleotide sequence of SEQ ID No 12 from the nucleotide at position 83 to the nucleotide at position 222; the nucleotide sequence of SEQ ID No 12 from the nucleotide at position 83 to the nucleotide at position 182; the nucleotide sequence of SEQ ID No 12 from

the nucleotide at position 83 to the nucleotide at position 132 ; and the nucleotide sequence of SEQ ID No 12 from the nucleotide at position 83 to the nucleotide at position 107) were used as nucleic acid fragments of insert for construction of chimeric genes capable of producing dsRNA.

5

This gene was chosen because its mutant allele has been reported in Arabidopsis to give distinct phenotypes. The CHS tt4(85) EMS mutant (Koornneef, 1990) produces inactive CHS resulting in no anthocyanin pigment in either the stem or seed-coat. Wildtype plants produce the purple-red pigment in both tissues.

10

In a first step, the respective fragments were PCR amplified using specific primers further comprising *attB1* and *attB2* recombination sites. *AttB1* and *attB2* specific primers were purchased from Life Technologies. The 25 and 50 bp fragments flanked by att sites were made by dimerization of the primers.

15

The following combinations of primers were used :

For the 400 bp fragment

Forward primer:

GGGGACAAGTTTGTACAAAAAAGCAGGCTGCACTGCTAACCCTGAGAACCATGTG

20 CTTC (SEQ ID No 14); and

Reverse primer:

GGGGACCACTTTGTACAAGAAAGCTGGGTCGCTTGACGGAAGGACGGAGACCAAG
AAGC (SEQ ID No 15).

25 For the 200 bp fragment

Forward primer:

GGGGACAAGTTTGTACAAAAAAGCAGGCTGCACTGCTAACCCTGAGAACCATGTG

CTTC (SEQ ID No 14); and

Reverse primer:

30 GGGGACCACTTTGTACAAGAAAGCTGGGTAGGAGCCATGTAAGCACACATGTGTG
GGTT (SEQ ID No 16).

For the 100 bp fragment

Forward primer:

- 5 GGGGACAAGTTTGTACAAAAAAGCAGGCTGCACTGCTAACCCTGAGAACCATGTG
CTTCAGGCGGAGTATCCTGACTACTACTTCCGCATCACCAACAGT (SEQ ID No 17);
and

Reverse primer:

- GGGGACCACTTTGTACAAGAAAGCTGGGTAACCTTCCTTGAGGTCGGTCATGTG
10 TTCACTGTTGGTGATGCGGAAGTAGTAGTCAGGATACTCCGCCTG (SEQ ID No 18).

For the 50 bp fragment

Forward primer:

- GGGGACAAGTTTGTACAAAAAAGCAGGCTGCACTGCTAACCCTGAGAACCATGTG
15 CTTCAGGCGGAGTATCCTGACTAC (SEQ ID No 19); and

Reverse primer:

GGGGACCACTTTGTACAAGAAAGCTGGGTGTAGTCAGGATACTCCGCCTGAAGCA
CATGGTTCTCAGGGTTAGCAGTGC (SEQ ID No 20).

- 20 For the 25 bp fragment

Forward primer:

GGGGACAAGTTTGTACAAAAAAGCAGGCTGCACTGCTAACCCTGAGAACCATGT
(SEQ ID No 21); and

Reverse primer:

- 25 GGGGACCACTTTGTACAAGAAAGCTGGGTACATGGTTCTCAGGGTTAGCAGTGC
(SEQ ID No 22).

- PCR amplification and recombination using the GATEWAY™ technology with the
commercially available BP Clonase (Life Technologies) were performed according to
30 the manufacturer's instructions (manual available on
<http://www.lifetech.com/content.cfm?pageid=2497>).

Bacterial colonies obtained on chloramphenicol-containing plates spread with *E. coli* DH5 α bacteria, transformed (by electroporation or by heatshocking RbCl₂ treated competent *E. coli* cells) with the *in vitro* recombination reaction were screened. Colonies containing the desired recombinant plasmid were obtained in each case. For the about 400 bp fragment 24 colonies were screened and 23 contained the desired construct with the 400 bp in inverted repeat, operably linked to the CaMV35S promoter. For the about 200 bp fragment 36 colonies were screened and 35 contained the desired construct with the 200 bp in inverted repeat, operably linked to the CaMV35S promoter. For the about 50 bp fragment 6 colonies were screened and 4 contained the desired construct with the 50 bp in inverted repeat, operably linked to the CaMV35S promoter. For the 25 bp fragment, 6 colonies were screened and 1 contained the desired construct with the 400 bp in inverted repeat, operably linked to the CaMV35S promoter. In a number of cases the structure was confirmed by sequence analysis.

These results show that this vector facilitates the rapid, efficient, and simple production of hpRNA (hairpin RNA constructs). pHELLSGATE is a T-DNA vector, with a high-copy-number origin of replication for ease of handling. Recombinant pHELLSGATE constructs can be directly transformed into *Agrobacterium* for transformation of the chimeric construct into plants. This system can be used in high throughput applications.

Example 3

Evaluation of plants comprising the chimeric genes of Example 2.

- The vectors containing the dsRNA producing chimeric constructs with the 400, 200, 100, 50 and 25 nucleotides of chalcone synthase in inverted repeat (Example 2) were introduced into *Agrobacterium tumefaciens* strain AGL1, GV3101 or LBA4404 either by electroporation or tri-parental mating.
- Transgenic Arabidopsis lines are obtained by transformation with these *Agrobacteria* using the dipping method of Clough and Bent (1998).

Chalcone synthase activity is monitored by visual observation of stem and leaf color (normally in plants grown under high light, and by unaided or microscope assisted visual observation of seed-coat color.

Most of the transgenic lines transformed with the above mentioned CHS silencing
5 constructs show pronounced silencing. The seed colour of most of these lines is virtually indistinguishable from seed of the *tt4(85)* mutant to the naked eye. Examination of the seed under a light microscope reveals that the degree of pigmentation is generally uniform in the cells of the coat of an individual seed, and among seeds of the same line.

10

Example 4

Construction of the acceptor vectors pHELLSGATE 4, pHELLSGATE 8, pHELLSGATE 11 and pHELLSGATE 12.

pHELLSGATE 4 was made by excising the DNA fragment comprising the *pdK* intron
15 and chloramphenicol resistance gene from pHELLSGATE (Example 1) with *HindIII* and *EcoRI* and replacing it with a *HindIII/EcoRI* DNA fragment containing only the *pdK* intron. The complete nucleotide sequence of pHELLSGATE 4 is represented in the sequence listing (SEQ ID No 23).

20 pHELLSGATE 8 was made by PCR amplification using pHellsgate DNA as a template and oligonucleotides with the sequence
5'GGGCTCGAGACAAGTTTGTACAAAAAAGCTG 3' and
5'GGCTCGAGACCACTTTGTACAAGAAAGC 3' as primers. These primers modify the attP sites within pHellsgate to attR sites. The resulting fragment was sequenced and
25 inserted into the *XhoI* site of a vector upstream of a DNA fragment containing the *pdK* intron fragment. Similarly an *XbaI/XbaI* fragment amplified with the oligonucleotides
5'GGGTCTAGACAAGTTTGTACAAAAAAGCTG 3' and 5'
GGGTCTAGACCACTTTGTACAAGAAAGC 3' as primers and pHELLSGATE as
template DNA to modify the attP sites of this cassette to attR sites. This fragment was
30 sequenced and inserted into the *XbaI* site of the intermediate described above downstream of the *pdK* intron. The complete nucleotide sequence of pHELLSGATE 8 is represented in the sequence listing (SEQ ID No 24) and a schematic figure can be found in Figure 4.

pHELLSGATE 11 is similar to pHELLSGATE 8 except that the pdk intron has been engineered to contain a branching point in the complementary strand such that splicing of the intron is independent of its orientation (a so-called "two-way intron"). The complete nucleotide sequence of pHELLSGATE 11 is represented in the sequence listing (SEQ ID No 25) and a schematic representation thereof can be found in Figure 4.

pHELLSGATE 12 is also similar to pHELLSGATE 8 except that the pdk intron has been duplicated as an inverted repeat. The complete nucleotide sequence of pHELLSGATE 12 is represented in the sequence listing (SEQ ID No 26) and a schematic representation thereof can be found in Figure 4.

15

Example 5

Use of the different pHELLSGATE vectors to generate dsRNA chimeric silencing genes targeted towards three different model target genes.

The efficiency in gene silencing of the different pHELLSGATE vectors was tested by inserting fragments of three target genes Flowering locus C (FLC) Ethylene insensitive 2 (EIN2) and Phytoene desaturase (PDC). For FLC a 390 bp fragment was used (from the nucleotide at position 303 to the nucleotide at position 692 of the nucleotide sequence available as Genbank Accession Nr AF116527). For EIN2 a 580 bp fragment was used (from the nucleotide at position 541 to the nucleotide at position 1120 of the nucleotide sequence available as Genbank Accession Nr AF141203). For PDS a 432 bp fragment was used (from the nucleotide at position 1027 to the nucleotide at position 1458 of the nucleotide sequence available as Genbank Accession Nr L16237). Genes of interest were amplified using gene specific primers with either a 5' attB1 extension (GGGGACAAGTTTGTACAAAAAAGCAGGCT) or an attB2 extension (GGGACCACTTTGTACAAGAAAGCTGGGT) using F1 Taq DNA polymerase (Fisher Biotec, Subiaco, WA, Australia) according to the manufacturer's protocol. PCR products were precipitated by adding 3 volumes TE and two volumes 30% (w/v) PEG 3000, 30mM MgCl₂ and centrifuging at 13000 g for 15 minutes. Recombination reaction of PCR products with either pDONR201 (Invitrogen, Groningen, The

Netherlands) or pHELLSGATE 4 were carried out in a total volume of 10 μ L with 2 μ L BP clonase buffer (Invitrogen), 1-2 μ L PCR product 150 ng plasmid vector and 2 μ L BP clonase (Invitrogen). The reaction was incubated at room temperature (25°C) for 1 h to overnight. After the incubation, 1 μ L proteinase K (2 μ g/ μ L; Invitrogen) was added and
5 incubated for 10 min at 37°C. 1-2 μ L of the mix was used to transform DH5 α , colonies were selected on the appropriate antibiotics. Clones were checked either by digestion of DNA minipreps or PCR. Recombination reactions from pDONR201 clones to pHellsgate 8, 11 or 12 were carried out in 10 μ L total volume with 2 μ L LR clonase buffer (Invitrogen), 2 μ L pDONR201 clone (approximately 150 ng), 300 ng pHellsgate
10 8, 11 or 12 and 2 μ L LR clonase (Invitrogen). The reaction was incubated overnight at room temperature, proteinase-treated and used to transform *E. coli* DH5 α as for the BP clonase reaction. Transformation of *Arabidopsis* was performed according to via the floral dip method (Clough and Bent, 1998). Plants were selected on agar solidified MS media supplemented with 100 mg/l timentin and 50 mg/l kanamycin. For *FLC* and
15 *PDS* constructs the C24 ecotype was used; for *EIN2* constructs Landsberg *erecta* was used. For scoring of *EIN2* phenotypes transformed T1 plants were transferred to MS media containing 50 μ M 1-aminocyclopropane-1-carboxylic acid (ACC) together with homozygous *EIN2*-silenced lines and wild type Landsberg *erecta* plants. T1 *FLC* hpRNA plants were scored by transferring to MS plates and scoring days to flower or
20 rosette leaves at flowering compared to C24 wild type plants and *flc* mutant lines. T1 *PDS* hpRNA plants were scored by looking at bleaching of the leaves. The results of the analysis of plants transformed with the different pHELLSGATE vectors are shown in Table 1.

25 All plants transformed with pHellsgate 4-*FLC* and pHellsgate 8-*FLC* flowered significantly earlier than wildtype C24 and in both cases plants flowering with the same number of rosette leaves as the *flc-20* line (carrying a stable Ds insertion in the first intron of the *FLC* gene) were observed. There was no clear difference in rosette leaves at flowering between the sets of plants transformed with the pHELLSGATE 4-
30 *FLC* and pHellsgate 8-*FLC* constructs.

A difference in the effectiveness of the pHELLSGATE 4-*EIN2* and pHELLSGATE 8-*EIN2* plants was observed. Of 36 transformants for pHG4-*EIN2* there were no plants

with an observable ACC-resistant phenotype under the conditions used for this experiment, whereas 8 of the 11 plants carrying the pHG8-EIN2 transgene showed some degree of ACC-resistance. The extent to which the pHG8-EIN2 plants were resistant to ACC was variable indicating that the severity of silencing varies between
5 transformants.

The great majority of plants carrying pHG4-PDS and pHG8-PDS showed a phenotype consistent with the loss of photoprotection due to the absence of carotenoids. The weakest phenotype was a bleaching of the cotyledons, with the true leaves not
10 bleaching at any stage in the life cycle. The bleached cotyledon phenotype was only seen in plants transformed with *PDS* hpRNA constructs; we confirmed that the plants with this phenotype also contained the *PDS* hpRNA construct (data not shown) strongly suggesting that this phenotype is due to *PDS* silencing and not bleaching from the kanamycin selection. Plants transformed with the pHELLSGATE 4-PDS
15 construct gave only this weak bleached cotyledon phenotype. In contrast the five of the pHELLSGATE 8-PDS plants had the weak phenotype and three showed a stronger phenotype with extensive or complete bleaching of the true leaves.

Table 1

| Construct | Test genes | T1 plants | Rate of silencing |
|---|--------------------|-----------|-------------------|
| HELLSGATE 4 | FLC | 13 | 12 |
| | EIN2 | 36 | 0 |
| | PDS | 12 | 11 |
| HELLSGATE 8 | FLC | 6 | 6 |
| | EIN2 | 11 | 8 |
| | PDS | 9 | 8 |
| HELLSGATE 11 | FLC | 2 | 2 |
| | EIN2 | 30 | 11 |
| | PDS | 11 | 11 |
| HELLSGATE 11 (intervening region in inverse orientation) | FLC EIN2 PDS | 8 | 6 |
| HELLSGATE 12 | FLC | 13 | 11 |
| | EIN2 | 28 | 12 |
| | PDS | | |
| HELLSGATE 12 (intervening region in inverse orientation) | FLC | 9 | 8 |
| | EIN2 | 5 | 2 |
| | PDS | | |
| | CHS | | |

References

- An *et al.* (1996) *The Plant Cell* 8, 15-30
- AzpiroLeehan and Feldmann (1997) *Trends Genet.* 13: 152-156
- 5 Clough and Bent (1998) *Plant J.* 16: 735-743
- Fire *et al.* (1998) *Nature* 391: 806-811
- Fromm *et al.* (1990) *Bio/Technology* 8: 833
- Gordon-Kamm *et al.* (1990) *The Plant Cell* 2: 603
- Hamilton *et al.* (1998) *Plant J.* 15: 737-746
- 10 Hoess *et al.* (1986) *Nucl. Acids Res.* 14: 2287
- Hudspeth *et al.* (1989) *Plant Mol. Biol.* 12: 579-589
- Keil *et al.* (1989) *EMBO J.* 8: 1323-1330
- Keller *et al.* (1988) *EMBO J.* 7: 3625-3633
- Keller *et al.* (1989) *Genes & Devel.* 3: 1639-1646
- 15 Koornneef (1990) *Theor. Appl. Gen.* 80: 852-857
- Landy (1993) *Current Opinions in Genetics and Development* 3: 699-707
- Landy (1989) *Ann. Rev. Biochem.* 58: 913
- Martienssen (1998) *Proc. Natl. Acad. Sci. USA* 95: 2021-2026
- Meyer *et al.* (1987) *Nature* 330: 677
- 20 Needleman and Wunsch (1970) *J. Mol. Biol.* 48: 443-453
- Peleman *et al.* (1989) *Gene* 85: 359-369
- Ross-MacDonald *et al.* (1999) *Nature* 402: 413-418
- Smith *et al.* (2000) *Nature* 407: 319-320
- Wagner and Sun (1998) *Nature* 391: 744-745
- 25 Waterhouse *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95: 13959-13964

Claims:

1. A vector comprising the following operably linked DNA fragments:
 - a) an origin of replication allowing replication in a recipient cell (1), preferably in bacteria; particularly in *Escherichia coli*.
 - 5 b) a selectable marker region (2) capable of being expressed in said recipient cell; and
 - c) a chimeric DNA construct comprising in sequence:
 - i) a promoter or promoter region (3) capable of being recognized by RNA polymerases of a eukaryotic cell;
 - 10 ii) a first recombination site (4), a second recombination site (5), a third recombination site (6) and a fourth recombination site (7);
 - iii) a 3' transcription terminating and polyadenylation region (8) functional in said eukaryotic cell;

wherein said first recombination site (4) and said fourth recombination site (7) are

 - 15 capable of reacting with a same recombination site, preferably are identical, and said second recombination site (5) and said third recombination site (6), are capable of reacting with a same recombination site, preferably are identical; and wherein said first recombination site (4) and said second recombination site (5) do not recombine with each other or with a same recombination site or said third
 - 20 recombination site (6) and said fourth recombination site (7) do not recombine with each other or with a same recombination site.
2. The vector of claim 1, wherein said first (4) and second recombination site (5) flank a second selectable marker gene (10) and said third (6) and fourth
 - 25 recombination site (7) flank a third selectable marker gene (9).
3. The vector of claim 1 or 2, wherein said chimeric DNA construct comprises a region flanked by intron processing signals (11), functional in said eukaryotic cell, located between said second recombination site (5) and said third recombination
 - 30 site (6).
4. The vector of claim 3, wherein said region flanked by intron processing signals is an intron sequence functional in said eukaryotic cell.

5. The vector of any one of claims 3 or 4, further comprising a fourth selectable marker gene (19), located between said second (5) and third recombination site (6).
- 5 6. The vector of any one of claims 1 to 5, wherein said selectable marker genes are selected from the group consisting of an antibiotic resistance gene, a tRNA gene, an auxotrophic marker, a toxic gene, a phenotypic marker, an antisense oligonucleotide; a restriction endonuclease; a restriction endonuclease cleavage site, an enzyme cleavage site, a protein binding site, an a sequence complementary
10 PCR primer.
7. The vector of any one of claims 1 to 6, wherein said promoter (3) is a plant-expressible promoter.
- 15 8. The vector of any one of claim 7, wherein said chimeric DNA construct is flanked by left and right border T-DNA sequences.
9. The vector of claim 8, further comprising a selectable marker gene capable of being expressed in plant cells located between said left and said right T-DNA border
20 sequences.
10. The vector of claim 8 or claim 9, further comprising an origin of replication capable of functioning in *Agrobacterium* sp.
- 25 11. The vector of any one of claims 1 to 10, wherein said first (4) and fourth recombination site (7) is attR1 comprising the nucleotide sequence of SEQ ID No 4 and said second (5) and third (6) recombination site is attR2 comprising the nucleotide sequence of SEQ ID No 5.
- 30 12. The vector of any one of claims 1 to 10, wherein said first (4) and fourth recombination site (7) is attP1 comprising the nucleotide sequence of SEQ ID No 10 and said second (5) and third (6) recombination site is attP2 comprising the nucleotide sequence of SEQ ID No 11.

13. A vector comprising the sequence of SEQ ID No 13.
14. A vector comprising the sequence of SEQ ID No 23.
- 5 15. A vector comprising the sequence of SEQ ID No 24.
16. A vector comprising the sequence of SEQ ID No 25.
- 10 17. A vector comprising the sequence of SEQ ID No 26.
18. A vector comprising the following operably linked DNA fragments:
- a) an origin of replication allowing replication in a recipient cell (1), preferably in bacteria; particularly in *Escherichia coli*.
 - 15 b) a selectable marker region (2) capable of being expressed in said recipient cell; and
 - c) a chimeric DNA construct comprising in sequence:
 - i) a promoter or promoter region (3) capable of being recognized by a prokaryotic RNA polymerase;
 - 20 ii) a first recombination site (4), a second recombination site (5), a third recombination site (6) and a fourth recombination site (7);
 - iii) a 3' transcription terminating and polyadenylation region (8) functional in said eukaryotic cell;
- wherein said first recombination site (4) and said fourth recombination site (7) are
- 25 capable of reacting with a same recombination site, preferably are identical, and said second recombination site (5) and said third recombination site (6), are capable of reacting with a same recombination site, preferably are identical; and wherein said first recombination site (4) and said second recombination site (5) do not recombine with each other or with a same recombination site or said third recombination site (6)
- 30 and said fourth recombination site (7) do not recombine with each other or with a same recombination site.

19. The vector of claim 18, wherein said RNA polymerase is a bacteriophage single subunit RNA polymerase.
20. A kit comprising the vector according to any one of claims 1 to 19.
- 5 21. The kit of claim 20, further comprising at least one recombination protein capable of recombining a DNA segment comprising at least one of said recombination sites.
22. A method for making a chimeric DNA construct capable of expressing a dsRNA in
10 a eukaryotic cell comprising the step of
a) combining in vitro:
i) a vector according to any one of claims 1 to 19;
ii) an insert DNA comprising a DNA segment of interest (12) flanked by
15 (1) a fifth recombination site (13) which is capable of recombining with
said first (4) or fourth recombination site (7) on said vector; and
(2) a sixth recombination site (14) which is capable of recombining with
said second (5) or third recombination site (6) on said vector;
iii) at least one site specific recombination protein capable of recombining said
first (4) or fourth (7) and said fifth recombination site (13) and said second
20 (5) or third (6) and said sixth recombination site (14);
b) allowing recombination to occur so as to produce a reaction mixture comprising
product DNA molecules, said product DNA molecule comprising in sequence:
i) said promoter or promoter region (3) capable of being recognized by RNA
polymerases of said eukaryotic cell;
25 ii) a recombination site (15) which is the recombination product of said first
(4) and said fifth recombination site (13);
iii) said DNA fragment of interest (12);
iv) a recombination site (16) which is the recombination product of said second
(4) and said sixth recombination site (14);
30 v) a recombination site (17) which is the recombination product of said third
(5) and said sixth recombination site (14);
vi) said DNA fragment of interest in opposite orientation (12);

- vii) a recombination site (18) which is the recombination product of said fourth (7) and said fifth recombination site (13); and
 - viii) said 3' transcription terminating and polyadenylation region (8) functional in said eukaryotic cell;
 - 5 c) selecting said product DNA molecules.
23. The method according to claim 22, wherein said selecting is carried out in vivo.
24. The method according to claim 22 or 23, wherein said insert DNA is a linear DNA
10 molecule.
25. The method according to claim 22 or 23, wherein said insert DNA is a circular DNA molecule.
- 15 26. The method according to any of claims 22 to 25, wherein said at least one recombination protein is selected from (i) Int and IHF and (ii) Int, Xis, and IHF.
27. The method according to any one of claims 22 to 25, wherein multiple insert DNAs comprising different DNA fragments of interest are processed
20 simultaneously.
28. A method for preparing a eukaryotic non-human organism wherein the phenotypic expression of a target nucleic acid of interest is reduced or inhibited, said method comprising:
- 25 a) preparing a chimeric DNA construct comprising a nucleic acid of interest (12) comprising a nucleotide sequence of at least 19 bp with at least 70% sequence identity to said target nucleic acid capable of expressing a dsRNA in cells of said eukaryotic non-human organism according to any one of the methods of claims 22 to 27;
- 30 b) introducing said chimeric DNA construct in cells of said eukaryotic non-human organism; and
- c) isolating said eukaryotic organism

29. The method of claim 28, wherein said eukaryotic organism is a plant.
30. A method for isolating a nucleic acid molecule involved in determining a particular trait
- 5 a) preparing a library of chimeric DNA constructs capable of expressing a dsRNA in cells of said eukaryotic non-human organism according to any one of the methods of claims 22 to 27;
- b) introducing individual representatives of said library of chimeric DNA constructs in cells of said eukaryotic non-human organism;
- 10 c) isolating a eukaryotic organism exhibiting said particular trait; and
- d) isolating said nucleic acid molecule.
31. The method according to claim 30, wherein said eukaryotic organism is a plant.
- 15 32. A eukaryotic non-human organism comprising a chimeric DNA construct obtainable through the methods of any one of claims 22 to 27.
33. The non-human eukaryotic organism according to claim 31 which is a plant.

1/6

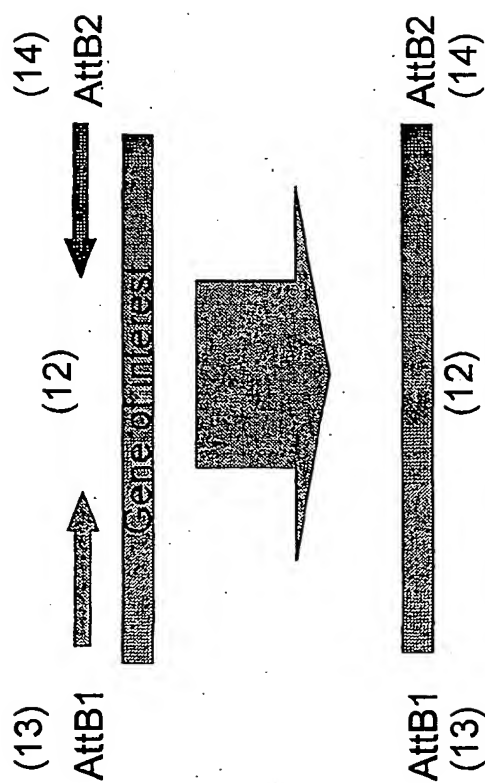


Figure 1A

2/6

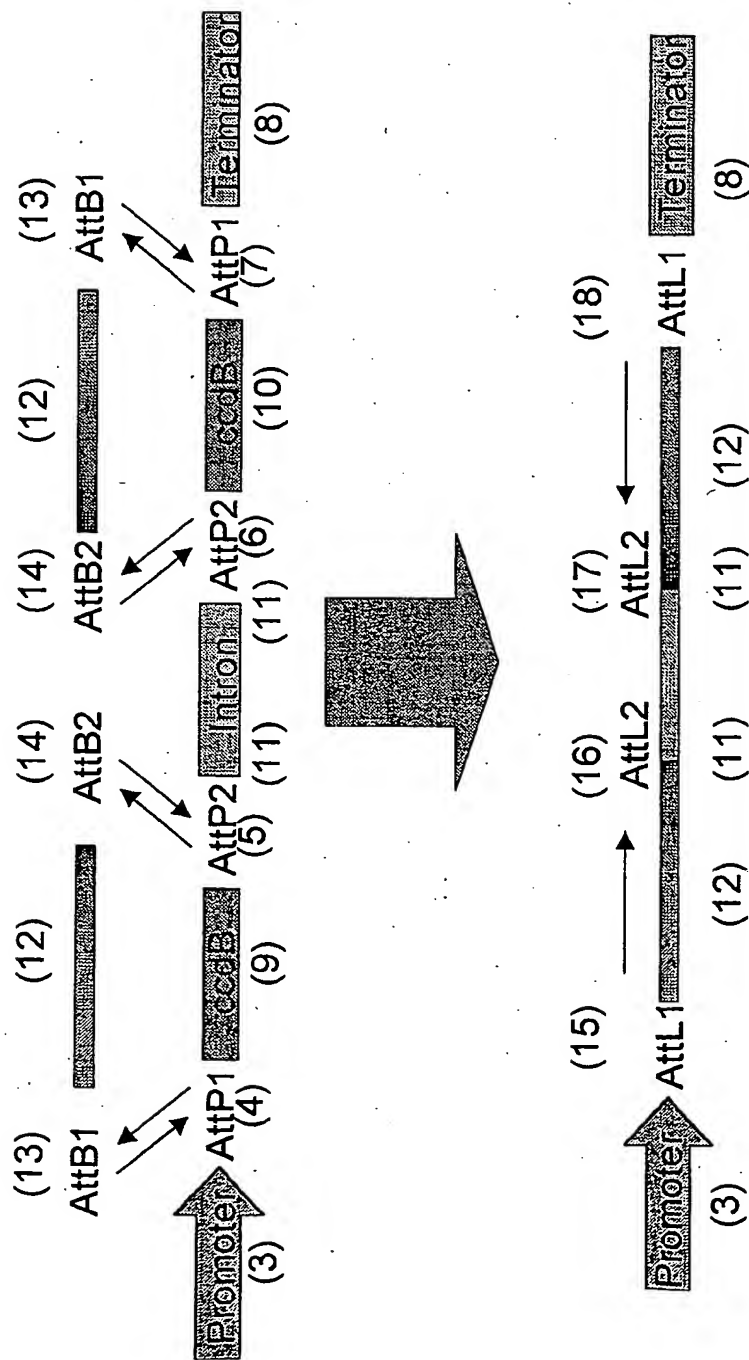


Figure 1B

3/6

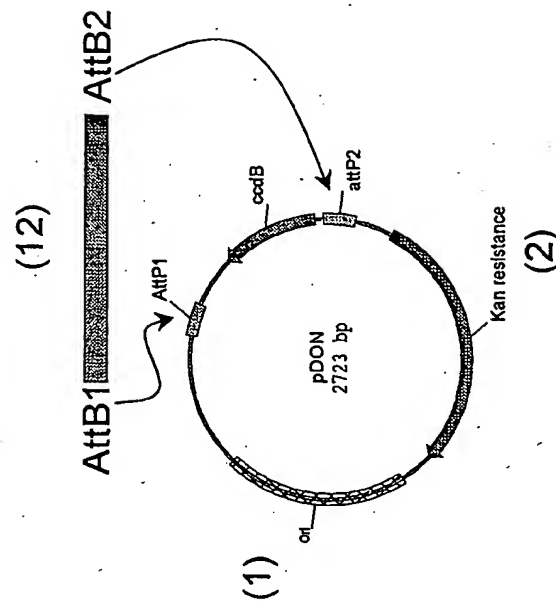


Figure 2B

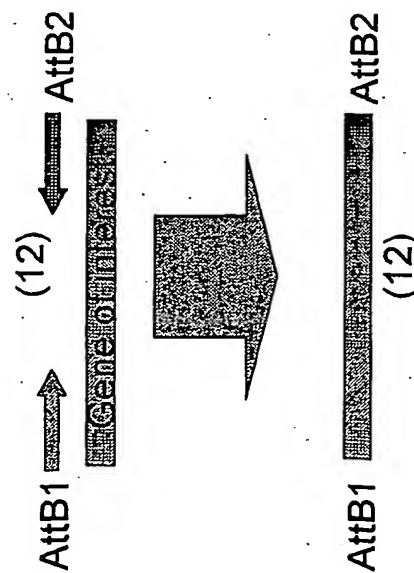


Figure 2A

4/6

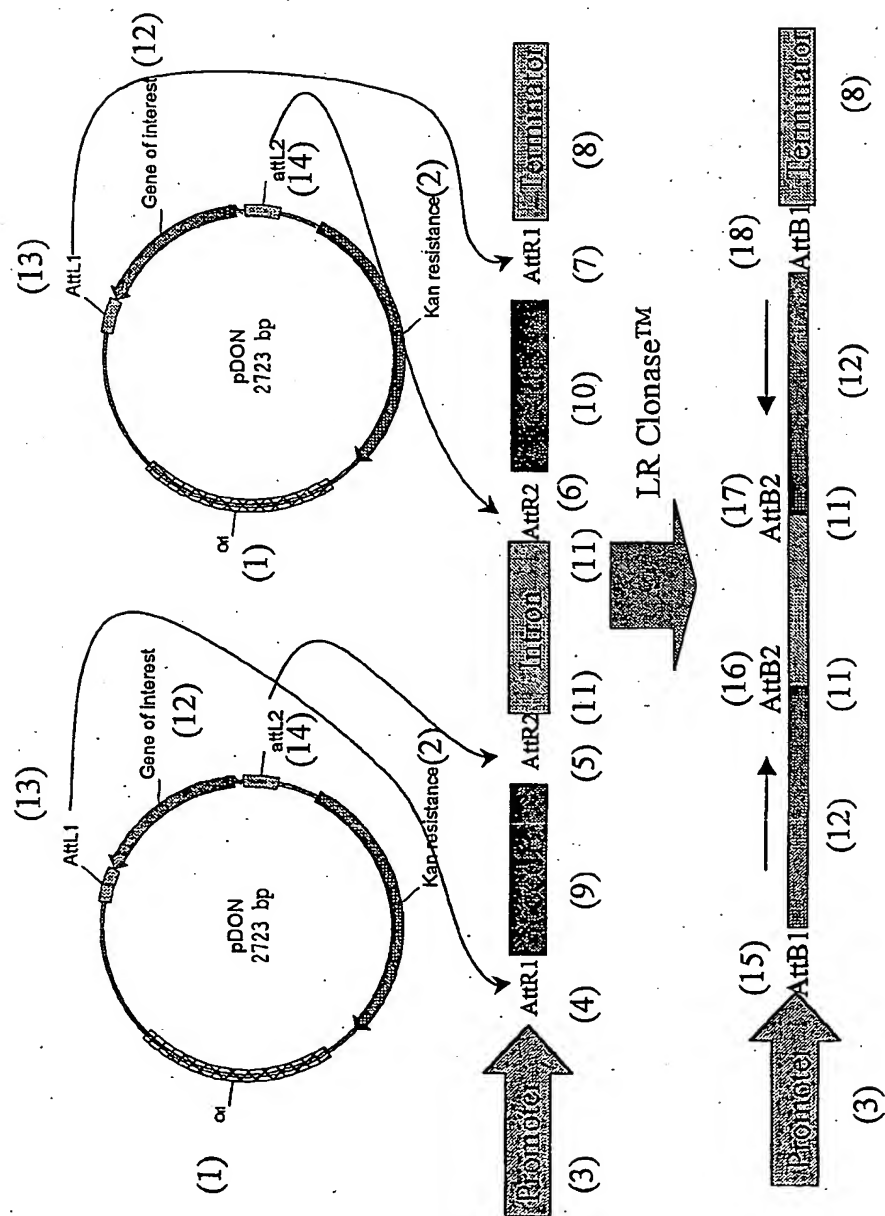


Figure 2C

5/6

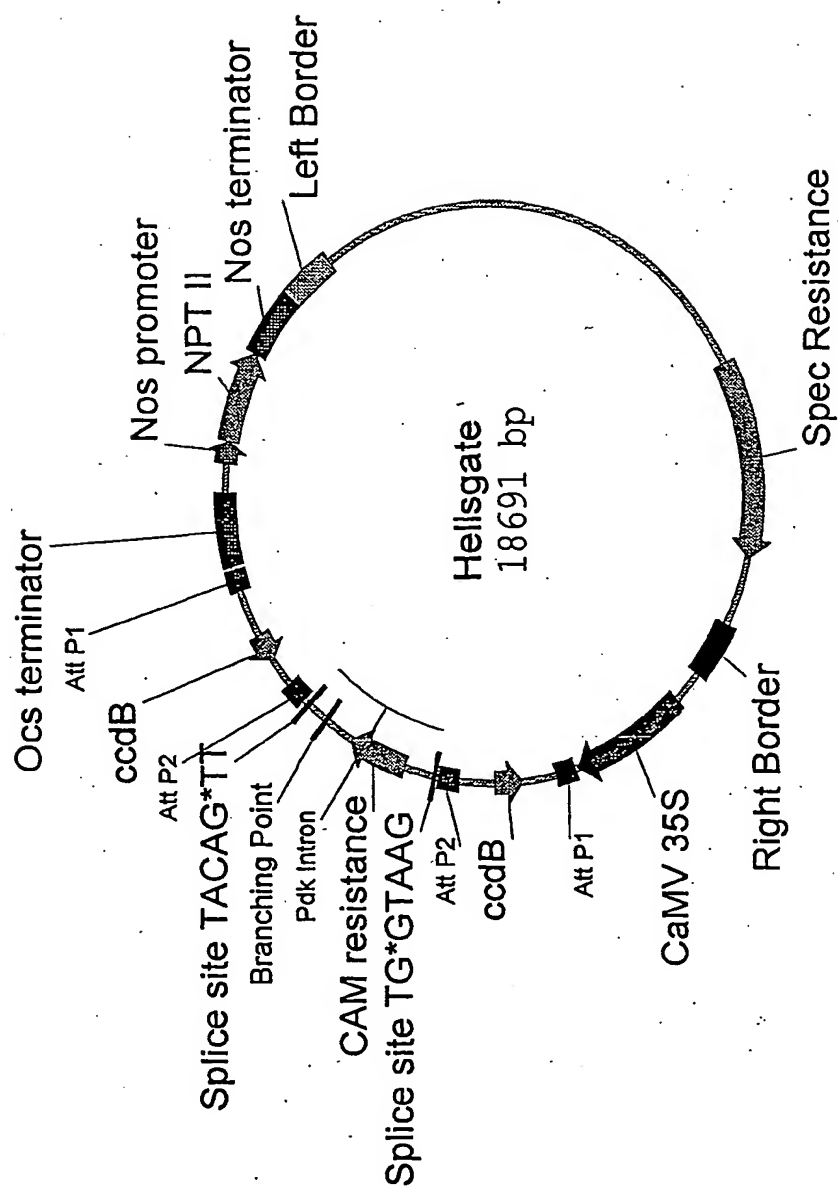


Figure 3

BEST AVAILABLE COPY

6/6

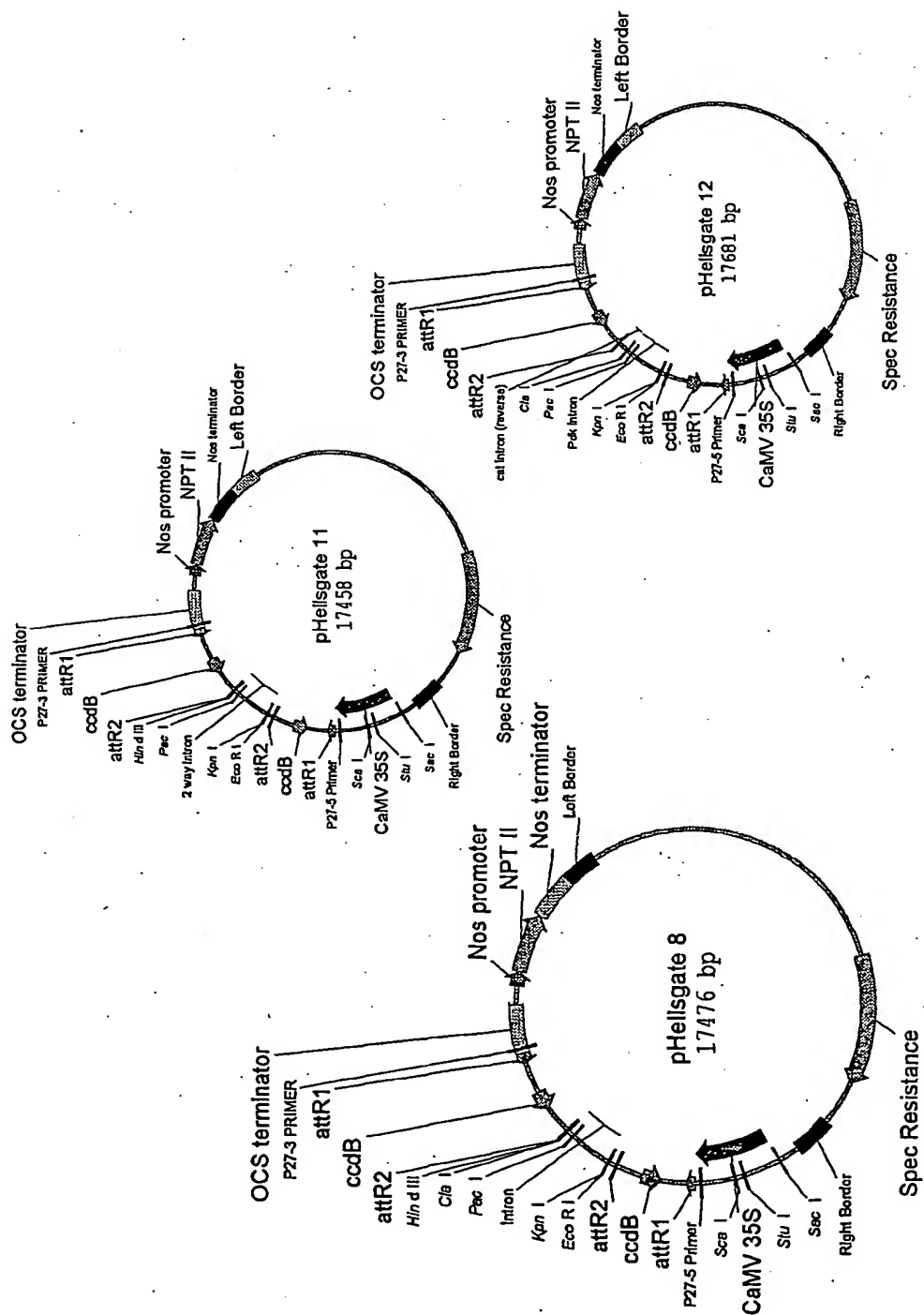


Figure 4

SEQUENCE LISTING

5 <110> COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION
<120> Method and means for producing efficient silencing constructs using
recombinational cloning
10 <130> 500255/MRO

<150> US60/264,067
<151> 2001-01-26
15 <150> US60/333,743
<151> 2001-11-29

<160> 26
<170> PatentIn version 3.1
20
<210> 1
<211> 25
<212> DNA
<213> Artificial sequence
25
<220>
<223> core sequence of recombination site attB1

<400> 1
30 agcctgcttt tttgtacaaa cttgt 25

<210> 2
<211> 25
35 <212> DNA
<213> Artificial sequence

<220>
<223> core sequence of recombination site attB2
40
<400> 2
agcctgcttt cttgtacaaa cttgt 25

<210> 3
<211> 25
45 <212> DNA
<213> Artificial sequence

<220>
<223> core sequence of recombination site attB3
50
<400> 3
accagcttt cttgtacaaa cttgt 25
55

<210> 4
<211> 25
60 <212> DNA
<213> Artificial sequence

BEST AVAILABLE COPY

<220>
<223> core sequence of recombination site attR1

5 <400> 4
gttcagcttt tttgtacaaa cttgt 25

<210> 5
10 <211> 25
<212> DNA
<213> Artificial sequence

<220>
15 <223> core sequence of recombination site attR2

<400> 5
gttcagcttt cttgtacaaa cttgt 25

20 <210> 6
<211> 25
<212> DNA
<213> Artificial sequence

25 <220>
<223> core sequence of recombination site attR3

<400> 6
30 gttcagcttt cttgtacaaa gttgg 25

<210> 7
<211> 25
35 <212> DNA
<213> Artificial sequence

<220>
<223> core sequence of recombination site attL1

40 <400> 7
agcctgcttt tttgtacaaa gttgg 25

45 <210> 8
<211> 25
<212> DNA
<213> Artificial sequence

50 <220>
<223> core sequence of recombination site attL2

<400> 8
55 agcctgcttt cttgtacaaa gttgg 25

<210> 9
<211> 25
<212> DNA
60 <213> Artificial sequence

<220>

<223> core sequence of recombination site attL3

<400> 9
 5 acccagcttt cttgtacaaa gttgg 25

<210> 10
 <211> 25
 <212> DNA
 10 <213> Artificial sequence

<220>
 <223> core sequence of recombination site attP1

15 <400> 10
 gttcagcttt tttgtacaaa gttgg 25

<210> 11
 20 <211> 25
 <212> DNA
 <213> Artificial sequence

<220>
 25 <223> core sequence of recombination site attP2,P3

<400> 11
 gttcagcttt cttgtacaaa gttgg 25

30 <210> 12
 <211> 1188
 <212> DNA
 <213> Artificial sequence

35 <220>
 <223> cDNA sequence of the Arabidopsis thaliana chalcone synthase coding region

40 <400> 12
 atggtgatgg ctggtgcttc ttctttggat gagatcagac aggctcagag agctgatgga 60
 cctgcaggca tcttggtat tggcactgct aacctgaga accatgtgct tcaggcggag 120
 45 tctctgact actacttccg catcaccaac agtgaacaca tgaccgacct caaggagaag 180
 ttcaagcgca tgtgcgacaa gtcgacaatt cggaacgct acatgcatct gacggaggaa 240
 ttcctcaagg aaaaccaca catgtgtgct tacatggctc cttctctgga caccagacag 300
 50 gacatcgtgg tggtcgaagt ccctaagcta ggcaaagaag cggcagtga ggccatcaag 360
 gagtggggcc agcccaagtc aaagatcact catgtcgtct tctgcactac ctccggcgctc 420
 55 gacatgcctg gtgctgacta ccagctcacc aagcttcttg gtctccgtcc ttccgtcaag 480
 cgtctcatga tgtaccagca aggttgcttc gccggcggt ctgtcctccg tatcgctaag 540
 gatctcgccg agaacaaccg tggagcacgt gtctcgttg tctgctctga gatcacagcc 600
 60 gttaccttcc gtggtccctc tgacaccac cttgactccc tcgtcgggtca ggctcttttc 660

agtgatggcg cgcgcgact cattgtgggg tcggaccctg acacatctgt cggagagaaa 720
cccatctttg agatgggtgc tgccgctcag accatccttc cagactctga tgggtgccata 780
5 gacggacatt tgagggaagt tgggtctcacc ttccatctcc tcaaggatgt tcccggcctc 840
atctccaaga acattgtgaa gagtctagac gaagcgttta aacctttggg gataagtgac 900
tggaactccc tcttctggat agcccaccct ggaggctcag cgatcctaga ccagggtggag 960
10 ataaagctag gactaaagga agagaagatg agggcgacac gtcacgtggt gagcgagtat 1020
ggaaacatgt cgagcgctg cgttctcttc atactagacg agatgaggag gaagtcagct 1080
15 aaggatgggt tggccacgac aggagaaggg ttggagtggg gtgtcttggt tggtttcgga 1140
ccagggtctca ctgttgagac agtcgtcttg cacagcgttc ctctctaa 1188

20 <210> 13
<211> 18691
<212> DNA
<213> Artificial sequence

25 <220>
<223> acceptor vector pHELLSGATE

<220>
<221> misc_feature
30 <222> (7922)..(9985)
<223> spectinomycin resistance

<220>
35 <221> misc_feature
<222> (10706)..(11324)
<223> right T-DNA border fragment

40 <220>
<221> misc_feature
<222> (11674)..(13019)
<223> CaMV35S promoter fragment

45 <220>
<221> misc_feature
<222> (17890)..(17659)
<223> attP1 recombination site (complement)

50 <220>
<221> misc_feature
<222> (17610)..(16855)
55 <223> ccdB selection marker (complement)

<220>
<221> misc_feature
60 <222> (16551)..(16319)
<223> attP2 recombination site (complement)

<220>
<221> misc_feature
<222> (14660)..(16258)
5 <223> pdk2 intron 2

<220>
<221> misc_feature
10 <222> (15002)..(15661)
<223> chloramphenicol resistance gene

<220>
15 <221> misc_feature
<222> (14387)..(14619)
<223> attP2 recombination site

<220>
20 <221> misc_feature
<222> (13675)..(13980)
<223> ccdB selection marker (complement)

<220>
25 <221> misc_feature
<222> (13048)..(13279)
<223> attP1 recombination site
30

<220>
<221> misc_feature
<222> (17922)..(18687)
35 <223> octopine synthase gene terminator region

<220>
<221> misc_feature
40 <222> (264)..(496)
<223> nopaline synthase gene promoter

<220>
45 <221> misc_feature
<222> (497)..(1442)
<223> nptII coding region

<220>
50 <221> misc_feature
<222> (1443)..(2148)
<223> nopaline synthase gene terminator

<220>
55 <221> misc_feature
<222> (2149)..(2706)
<223> a left T-DNA border region
60

<400> 13

| | | |
|----|--|------|
| | ggccgcacta gtgatatccc gcggccatgg cggccgggag catgcgacgt cgggccaat | 60 |
| | tcgcctata gtgagtcgta ttacaattca ctggcgcgctg ttttacaacg tcgtgactgg | 120 |
| 5 | gaaaaccctg gcgttaccga acttaatcgc cttgcagcac atccccctt cgccagctgg | 180 |
| | cgtaatagcg aagaggcccg caccgatcgc ccttccaac agttgcgcag cctgaatggc | 240 |
| | gaatggaaat tgtaaacgtt aatgggttct tggagttaa tgagctaagc acatacgtca | 300 |
| 10 | gaaaccatta ttgcgcgttc aaaagtcgcc taaggtcact atcagctagc aaatatttct | 360 |
| | tgtcaaaaat gctccactga cgttcataa attcccctcg gtatccaatt agagtctcat | 420 |
| 15 | attcactctc aatccaaata atctgcaatg gcaattacct tatccgcaac ttctttacct | 480 |
| | atttcgccc ggatccgggc aggttctccg gccgcttggg tggagaggct attcggctat | 540 |
| | gactgggcac aacagacaat cggtctctct gatgccgcg tgttcgggct gtcagcgag | 600 |
| 20 | gggcgcgcgg ttctttttgt caagaccgac ctgtccggtg ccctgaatga actgcaggac | 660 |
| | gaggcagcgc ggctatcgtg gctggccacg acgggcgttc cttgcgcagc tgtgctcgac | 720 |
| 25 | gttgtcactg aagcgggaag ggactggctg ctattggcg aagtgcggg gcaggatctc | 780 |
| | ctgtcatctc accttgctcc tgccgagaaa gtatccatca tggctgatgc aatgcggcgg | 840 |
| | ctgcatacgc ttgatccggc tacctgccca ttcgaccacc aagcgaaaca tcgcatcgag | 900 |
| 30 | cgagcacgta ctcgatgga agccggtctt gtcgatcagg atgatctgga cgaagagcat | 960 |
| | caggggctcg cgccagccga actgttcgcc aggtcaagg cgcgcatgcc cgacggcgag | 1020 |
| 35 | gatctcgtcg tgaccatgg cgatgcctgc ttgccgaata tcatggtgga aaatggccgc | 1080 |
| | ttttctggat tcatcgactg tggccggctg ggtgtggcg accgctatca ggacatagcg | 1140 |
| | ttggctaccc gtgatattgc tgaagagctt ggcggcgaat gggctgaccg cttcctcgtg | 1200 |
| 40 | ctttacggta tcgcgcctcc cgattcgcag cgcatcgct tctatcgct tcttgacgag | 1260 |
| | ttcttctgag cgggactctg gggttcgaaa tgaccgacca agcgacgccc aacctgccat | 1320 |
| 45 | cacgagattt cgattccacc gccgccttct atgaaagggt gggcttcgga atcgttttcc | 1380 |
| | gggacgcgg ctggatgatc ctccagcgcg gggatctcat gctggagttc ttcgccacc | 1440 |
| | ccgatccaac acttacgttt gcaacgtcca agagcaaata gaccacgaac gccggaagg | 1500 |
| 50 | tgccgcagcg tgtggattgc gtctcaattc tctcttgag gaatgcaatg atgaatatga | 1560 |
| | tactgactat gaaactttga gggaatactg cctagaccg tcacctcata acgtgcatca | 1620 |
| 55 | tgcatgcctt gacaacatgg aacatcgcta ttttctgaa gaattatgct cgttgaggga | 1680 |
| | tgtcgcggca attgcagcta ttgccaatc cgaactaccc ctcacgcatg cattcatcaa | 1740 |
| | tattattcat gcggggaaag gcaagattaa tccaactggc aaatcatcca gcgtgattgg | 1800 |
| 60 | taacttcagt tccagcgact tgattcgttt tgggtgctacc cacgttttca ataaggacga | 1860 |

gatggtggag taaagaagga gtgcgtcgaa gcagatcggt caaacatttg gcaataaagt 1920
 ttcttaagat tgaatcctgt tgccggtcct gcgatgatta tcatataatt tctgttgaat 1980
 5 tacgttaagc atgtaataat taacatgtaa tgcatgacgt tatttatgag atgggttttt 2040
 atgattagag tcccgcaatt atacatttaa tacgcgatag aaaacaaaat atagcgcgca 2100
 aactaggata aattatcgcg cgcggtgtca tctatgttac tagatcgaat taattccagg 2160
 10 cgggtgaagg caatcagctg ttgccgtct cactggtgaa aagaaaaacc accccagtac 2220
 attaaaaacg tccgcaatgt gttattaagt tgtctaagcg tcaatttggt tacaccacaa 2280
 15 tatatcctgc caccagccag ccaacagctc cccgaccggc agctcggcac aaaatcacca 2340
 ctcgatacag gcagcccatc agtccgggac ggcgtcagcg ggagagccgt tgtaaggcgg 2400
 cagactttgc tcatgttacc gatgctattc ggaagaacgg caactaagct gccgggtttg 2460
 20 aaacacggat gatctcgcgg agggtagcat gttgattgta acgatgacag agcgttgctg 2520
 cctgtgatca aatatcatct ccctcgcaga gatccgaatt atcagccttc ttattcattt 2580
 25 ctgcgttaac cgtgacaggc tgtcgatctt gagaactatg ccgacataat aggaaatcgc 2640
 tggataaagc cgctgaggaa gctgagtggc gctatttctt tagaagtga cgttgacgat 2700
 gtcgacggat cttttccgct gcataaccct gcttcggggg cattatagcg attttttcgg 2760
 30 tatatccatc ctttttcgca cgatatacag gattttgcca aagggttcgt gtagactttc 2820
 cttggtgtat ccaacggcgt cagccgggca ggatagggtga agtaggccca cccgcgagcg 2880
 35 ggtgttcctt cttcactgtc ctttattcgc acctggcggg gctcaacggg aatcctgctc 2940
 tgcgaggctg gccggctacc gccggcgtaa cagatgaggg caagcggatg gctgatgaaa 3000
 ccaagccaac caggggtgat gctgccact tactgattta gtgtatgatg gtgtttttga 3060
 40 ggtgtctcag tggcttctgt ttctatcagc tgcctcctc gttcagctac tgacgggggtg 3120
 gtgcgtaacg gcaaaagcac cgccggacat cagcgctatc tctgctctca ctgccgtaaa 3180
 45 acatggcaac tgcagttcac ttacaccgct tctcaaccg gtacgcacca gaaaatcatt 3240
 gatatggcca tgaatggcgt tggatgccg gcaacagccc gcattatggg cgttgccctc 3300
 aacacgattt tacgtcactt aaaaaactca ggccgcagtc ggtaacctcg cgcatacagc 3360
 50 cgggcagtga cgtcatcgtc tgcgcggaaa tggacgaaca gtggggctat gtcggggcta 3420
 aatcgcgccca gcgctggctg ttttacgctg atgacagtct ccggaagacg gttgttgccg 3480
 55 acgtattcgg tgaacgact atggcgacgc tggggcgtct tatgagcctg ctgtcaccct 3540
 ttgacgtggg gatatggatg acggatggct ggccgctgta tgaatcccgc ctgaagggaa 3600
 agctgcacgt aatcagcaag cgatatacgc agcgaattga gcgcataac ctgaatctga 3660
 60 ggcagcacct ggcacggctg ggacggaagt cgctgtcgtt ctcaaatcg gtggagctgc 3720

atgacaaagt catcgggcat tatctgaaca taaaacacta tcaataagtt ggagtcatta 3780
cccaaccagg aagggcagcc cacctatcaa ggtgtactgc cttccagacg aacgaagagc 3840
5 gattgaggaa aaggcggcgg cggccggcat gaggctgtcg gcctacctgc tggccgtcgg 3900
ccagggttac aaaatcacgg gcgtcgtgga ctatgagcac gtccgcgagc tggcccgcac 3960
10 caatggcgac ctgggcccgc tggggggcct gctgaaactc tggctcaccg acgaccgcg 4020
cacggcgagg ttcggtgatg ccacgatcct cgccctgtcg gcgaagatcg aagagaagca 4080
ggacgagctt ggcaaggcca tgatgggcgt ggtccgcccg agggcagagc catgactttt 4140
15 ttagccgcta aaacggccgg ggggtgcgcg tgattgcaa gcacgtccc atgcgtcca 4200
tcaagaagag cgacttcgag gagctggtat tcgtgcaggg caagattcgg aataccaagt 4260
acgagaagga cggccagacg gtctacggga ccgacttcac tgccgataag gtggattatc 4320
20 tggacaccaa ggcaccaggc ggggtcaaactc aggaataagg gcacattgcc ccggcgtgag 4380
tcggggcaat cccgcaagga ggggtgaatga atcggacgtt tgaccggaag gcatacaggc 4440
25 aagaactgat cgacgcgggg ttttccgccc aggatgccga aaccatcgca agccgcaccg 4500
tcatgcgtgc gcccgcgaa accttcacgt ccgtcggctc gatggtccag caagctacgg 4560
ccaagatcga gcgcgacagc gtgcaactgg ctccccctgc cctgcccgcg ccacgcggccg 4620
30 ccgtggagcg ttcgcgtcgt ctcgacagc aggcggcagg tttggcgaag tcgatgacca 4680
tcgacacgag aggaactatg acgaccaaga agcgaaaaac cgccggcgag gacctggcaa 4740
35 aacaggtcag cgaggccaag caggccgcgt tgctgaaaca cacgaagcag cagatcaagg 4800
aatgcagct ttccttggtc gatattgcgc cgtggccgga cacgatgcga gcgatgcaa 4860
acgacacggc ccgctctgcc ctgttcacca cgcgcaaca gaaaatcccg cgcgaggcgc 4920
40 tgcaaaaaca ggtcattttc cacgtcaaca aggacgtgaa gatcacctac accggcgtcg 4980
agctgcgggc cgacgatgac gaactggtgt ggcagcagg gttggagtac gcgaagcgca 5040
45 cccctatcgg cgagccgatc accttcacgt tctacgagct ttgccaggac ctgggctggt 5100
cgatcaatgg ccggtattac acgaaggccg aggaatgcct gtcgcgccta caggcgacgg 5160
cgatgggctt cacgtccgac cgcgttgggc acctggaatc ggtgtcgtcg ctgcaccgct 5220
50 tccgcgtcct ggaccgtggc aagaaaacgt ccggttgcca ggtcctgatc gacgaggaaa 5280
tcgtcgtgct gtttgctggc gaccactaca cgaaattcat atgggagaag taccgcaagc 5340
55 tgctgcggac gggccgacgg atgttcgact atttcagctc gcaccgggag ccgtaccgcg 5400
tcaagctgga aaccttcgc ctcatgtgcg gatcggtatc caccgcgtg aagaagtggc 5460
gcgagcaggt cggcgaagcc tgcaagagt tgcgaggcag cggcctggtg gaacacgcct 5520
60 gggccaatga tgacctggtg cattgcaaac gctagggcct tgtggggtca gttccggctg 5580

ggggttcagc agccagcgct ttactggcat ttcaggaaca agcggggcact gctcgacgca 5640
 cttgcttcgc tcagtatcgc tcgggacgca cggcgcgctc tacgaactgc cgataaacag 5700
 5 aggattaaaa ttgacaattg tgattaaggc tcagattcga cggcttggag cggccgacgt 5760
 gcaggatttc cgcgagatcc gattgtcggc cctgaagaaa gctccagaga tgttcgggtc 5820
 10 cgtttacgag cagcaggaga aaaagcccat ggaggcgttc gctgaacggg tgcgagatgc 5880
 cgtggcattc ggccctaca tcgacggcga gatcattggg ctgtcggctc tcaaacagga 5940
 ggacggcccc aaggacgctc acaaggcgca tctgtccggc gttttcgtgg agcccgaaca 6000
 15 gcgaggccga ggggtcgcgc gtatgtgctc gcgggcgttg ccggcgggtt tattgtcgt 6060
 gatgatcgtc cgacagattc caacgggaat ctggtggatg cgcattctca tcctcggcgc 6120
 acttaatat tcgctattct ggagcttgtt gtttatttcg gtctaccgcc tgccgggcgg 6180
 20 ggtcgcggcg acggtaggcg ctgtgcagcc gctgatggc gtgttcattc ctgccgtctc 6240
 gctaggtagc ccgatacgat tgatggcggc cctgggggct atttgcgga ctgccggcgt 6300
 25 ggcgctgttg gtgttgacac caaacgcagc gctagatcct gtcggcgctc cagcgggcct 6360
 ggcgggggcg gtttccatgg cgttcggaac cgtgctgacc cgcaagtggc aacctccgt 6420
 gcctctgctc acctttaccg cctggcaact ggcggccgga ggacttctgc tcgttccagt 6480
 30 agcttttagtg tttgatccgc caatccgat gcctacagga accaatgttc tcggcctggc 6540
 gtggctcggc ctgatcggag cgggtttaac ctacttctt tggttccggg ggatctcgcg 6600
 35 actcgaacct acagtgttt cttactggg ctttctcagc cgggatggcg ctaagaagct 6660
 attgccgcgc atcttcatat gcggtgtgaa ataccgcaca gatgcgtaag gagaaaatac 6720
 cgcacaggc gctcttccgc ttctcgcgc actgactcgc tgcgctcggc cgttcggctg 6780
 40 cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg tatccacaga atcaggggat 6840
 aacgcaggaa agaactgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc 6900
 45 gcgttgctgg cgtttttcca taggtccgc cccctgacg agcatcaca aaatcgacgc 6960
 tcaagtcaga ggtggcga aa cccgacagga ctataaagat accaggcgtt tccccctgga 7020
 agctccctcg tgcgctctcc tgttcggacc ctgccgtta ccggatacct gtccgccttt 7080
 50 ctcccttcgg gaagcgtggc gctttctcaa tgctcagct gtaggtatct cagttcggtg 7140
 taggtcgttc gctccaagct gggctgtgtg cacgaacccc cgttcagcc cgaccgctgc 7200
 55 gccttatccg gtaactatcg tcttgagtcc aaccggtaa gacacgactt atcgccactg 7260
 gcagcagcca ctggaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc 7320
 ttgaagtggg ggcctaacta cggctacact agaaggacag ttttggtat ctgcgctctg 7380
 60 ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc 7440

| | | | | | | | |
|----|------------|------------|------------|------------|-------------|------------|------|
| | gctggtagcg | gtgggttttt | tgtttgcaag | cagcagatta | cgcgagaaa | aaaaggatat | 7500 |
| | caagaagatc | ctttgatctt | ttctacgggg | tctgacgctc | agtggaacga | aaactcacgt | 7560 |
| 5 | taagggattt | tggtcatgag | attatcaaaa | aggatcttca | cctagatcct | tttaaattaa | 7620 |
| | aaatgaagtt | ttaaataaat | ctaaagtata | tatgagtaaa | cttgggtctga | cagttacca | 7680 |
| | tgcttaatac | gtgaggcacc | tatctcagcg | atctgtctat | ttcgttcac | catagttgcc | 7740 |
| 10 | tgactccccg | tcgtgtagat | aactacgata | cgggagggct | taccatctgg | ccccagtgtc | 7800 |
| | gcaatgatac | cgcgagaccc | acgctcaccg | gctccagatt | tatcagcaat | aaaccagcca | 7860 |
| 15 | gccggaaggg | ccgagcgcag | aagtggctct | gcaactttat | ccgcctccat | ccagtctatt | 7920 |
| | aaacaagtgg | cagcaacgga | ttcgcaaac | tgctcagcct | tttgtgcaa | aagccgcgcc | 7980 |
| | aggtttgcca | tccgtgtg | caggcgtag | gcgtcatatg | aagatttcgg | tgatccctga | 8040 |
| 20 | gcaggtggcg | gaaacattgg | atgctgagaa | ccatttcatt | gttcgtgaag | tgttcgatgt | 8100 |
| | gcacctatcc | gaccaaggct | ttgaactatc | taccagaagt | gtgagcccct | accggaagga | 8160 |
| 25 | ttacatctcg | gatgatgact | ctgatgaaga | ctctgcttgc | tatggcgcat | tcacgacca | 8220 |
| | agagcttgct | gggaagattg | aactcaactc | aacatggaac | gatctagcct | ctatcgaaca | 8280 |
| | cattgttggt | tcgcacacgc | accgaggcaa | aggagtcg | cacagtctca | tcgaatttgc | 8340 |
| 30 | gaaaaagtgg | gactaagca | gacagctcct | tggcatacga | ttagagacac | aaacgaacaa | 8400 |
| | tgtacctgcc | tgcaatttgt | acgcaaaatg | tggtttact | ctcgccggca | ttgacctgtt | 8460 |
| 35 | cacgtataaa | actagacctc | aagtctcgaa | cgaaacagcg | atgtactgg | actggttctc | 8520 |
| | gggagcacag | gatgacgcct | aacaattcat | tcaagccgac | accgcttcgc | ggcgccgctt | 8580 |
| | aattcaggag | ttaaacaatc | tgagggaagc | ggtgatcgcc | gaagtatcga | ctcaactatc | 8640 |
| 40 | agaggtagtt | ggcgatcatc | agcgccatct | cgaaccgacg | ttgctggccg | tacatttgta | 8700 |
| | cggctccgca | gtggatggcg | gcctgaagcc | acacagtgat | attgatttgc | tggttacgg | 8760 |
| 45 | gaccgtaagg | cttgatgaaa | caacgcggcg | agctttgatc | aacgaccttt | tggaaacttc | 8820 |
| | ggcttccctt | ggagagagcg | agattctccg | cgctgtagaa | gtcaccattg | ttgtgcacga | 8880 |
| | cgacatcatt | ccgtggcggt | atccagctaa | gcgcgaactg | caatttggag | aatggcagcg | 8940 |
| 50 | caatgacatt | cttgacagga | tcttcgagcc | agccacgac | gacattgatc | tggtatctt | 9000 |
| | gctgacaaaa | gcaagagaac | atagcgttgc | cttggtaggt | ccagcggcgg | aggaactctt | 9060 |
| 55 | tgatccgggt | cctgaacagg | atctatttga | ggcgctaaat | gaaaccttaa | cgctatggaa | 9120 |
| | ctcgccgccc | gactgggctg | gcgatgagcg | aaatgtagtg | cttacgttgt | cccgcatttg | 9180 |
| | gtacagcgca | gtaaccggca | aaatcgcgcc | gaaggatgtc | gctgccgact | gggcaatgga | 9240 |
| 60 | gcgcctgccg | gcccagtatc | agcccgtcat | acttgaagct | aggcaggctt | atcttgga | 9300 |

agaagatcgc ttggcctcgc gcgcagatca gttggaagaa tttgttcaact acgtgaaagg 9360
cgagatcacc aaggtagtcg gcaaataatg tctaacaatt cgttcaagcc gacgccgctt 9420
5 cgcggcgcgg cttaactcaa gcgttagaga gctggggaag actatgcgcg atctgttgaa 9480
ggtggttcta agcctcgtac ttgcgatggc atcggggcag gcaattgctg acctgccaat 9540
10 tgttttagtg gatgaagctc gtcttccta tgactactcc ccaccaact acgacatttc 9600
tccaagcaac tacgacaact ccataagcaa ttacgacaat agtccatcaa attacgacaa 9660
ctctgagagc aactacgata atagttcatc caattacgac aatagtcgca acggaaatcg 9720
15 taggcttata tatagcgcaa atgggtctcg cactttcgcc ggctactacg tcattgccaa 9780
caatgggaca acgaacttct tttccacatc tggcaaaagg atgttctaca ccccaaaagg 9840
ggggcgcggc gtctatggcg gcaaagatgg gagcttctgc ggggcattgg tcgtcataaa 9900
20 tggccaattt tcgcttgccc tgacagataa cggcctgaag atcatgtatc taagcaacta 9960
gcctgctctc taataaaatg ttaggagctt ggctgccatt tttgggggtga ggccgttcgc 10020
25 ggccgagggg cgcagcccct ggggggatgg gaggcccgcg ttagcgggcc gggaggggttc 10080
gagaaggggg ggcaccccc ttccggctgc gcggtcacgc gccagggcgc agccctggtt 10140
aaaaacaagg ttataaaata ttggtttaa agcaggtaa aagacagggt agcgggtggcc 10200
30 gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 10260
tgtcaatagg tgcccccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 10320
35 ccctcatctg tcagtagtcg cgcacctcaa gtgtcaatac cgcagggcac ttatccccag 10380
gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcca 10440
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 10500
40 gccgggtgag tcggcccctc aagtgtcaac gtccgcccct catctgtcag tgagggccaa 10560
gttttcgcgc aggtatccac aacgcgcgcg gccggccgcg gtgtctcgca cacggcttcg 10620
45 acggcgtttc tggcgcttt gcagggccat agacggccgc cagcccagcg gcgagggcaa 10680
ccagcccgtg gagcgtcga aagggtcgac atcttgctgc gttcggatat tttcgtggag 10740
ttcccgcac agaccggat tgaaggcgag atccagcaac tcgcgccaga tcactctgtg 10800
50 acggaacttt ggcgcgtgat gactggccag gacgtcggcc gaaagagcga caagcagatc 10860
acgattttcg acagcgtcgg atttgcgac gaggattttt cggcgctgcg ctacgtccgc 10920
55 gaccgcgttg agggatcaag ccacagcagc cactcgacc ttctagccga cccagacgag 10980
ccaagggatc tttttggaat gctgctccgt cgtcaggctt tccgacgttt ggggtggtga 11040
acagaagtca ttatcgtacg gaatgccagc actcccagg ggaacctgt ggttgcatg 11100
60 cacatacaaa tggacgaacg gataaacctt ttcacgccct tttaaatata cgttattcta 11160

ataaacgctc ttttctctta ggtttaccgc ccaatatatc ctgtcaaaca ctgatagttt 11220
aaactgaagg cgaggaaacga caatctgacg atgagcggag aattaaggga gtcacgttat 11280
5 gacccccgcc gatgacgcgg gacaagccgt tttacgtttg gaactgacag aaccgcaacg 11340
attgaaggag ccaactcagcc ccaatacgca aaccgcctct ccccgcgcggt tggccgattc 11400
attaatgcag ctggcacgac aggtttcccg actggaaagc gggcagtgag cgcaacgcaa 11460
10 ttaatgtgag ttagctcact cattaggcac ccaggcttt acactttatg cttccggctc 11520
gtatgttgtg tgggaattgtg agcggataac aatttcacac aggaaacagc tatgaccatg 11580
15 attacgcaa gctatttagg tgacactata gaatactcaa gctatgcac caacgcgttg 11640
ggagctctcc catatcgacc tgcaggcggc cgctcgacga attaattcca atccacaaa 11700
aatctgagct taacagcaca gttgctctc tcagagcaga atcgggtatt caacaccctc 11760
20 atatcaacta ctacgttgtg tataacggtc cacatgccgg tatatacgat gactgggggt 11820
gtacaaggc ggcaacaaac ggcgttcccg gagttgcaca caagaaattt gccactatta 11880
25 cagaggcaag agcagcagct gacgcgtaca caacaagtca gcaaacagac aggttgaact 11940
tcattcccaa aggagaagct caactcaagc ccaagagctt tgctaaggcc ctaacaagcc 12000
caccaaagca aaaagccac tggctcacgc taggaaccaa aaggccagc agtgatccag 12060
30 ccccaaaaga gatctcctt gccccggaga ttacaatgga cgatttcctc tatctttacg 12120
atctaggaag gaagtgcga ggtgaagggt acgacactat gttcaccact gataatgaga 12180
35 aggttagcct cttcaatttc agaaagaatg ctgaccaca gatggttaga gaggcctacg 12240
cagcaggtct catcaagacg atctaccga gtaacaatct ccaggagatc aaataccttc 12300
ccaagaagg taaagatgca gtcaaaagat tcaggactaa ttgcatcaag aacacagaga 12360
40 aagacatatt tctcaagatc agaagtacta ttccagtatg gacgattcaa ggttgcttc 12420
ataaaccaag gcaagtaata gagattggag tctctaaaaa ggtagttcct actgaatcta 12480
45 aggccatgca tggagtctaa gattcaaact gaggatctaa cagaactcgc cgtgaagact 12540
ggcgaacagt tcatacagag tcttttacga ctcaatgaca agaagaaaat cttcgtcaac 12600
atggtggagc acgacactct ggtctactcc aaaaatgtca aagatacagt ctcagaagac 12660
50 caaagggcta ttgagacttt tcaacaaagg ataatttcgg gaaacctcct cggattccat 12720
tgcccagcta tctgtcactt catcgaaagg acagtagaaa aggaagggtg ctctacaaa 12780
55 tgccatcatt gcgataaagg aaaggctatc attcaagatc tctctgccga cagtgggtccc 12840
aaagatggac cccacccac gaggaacatc gtggaaaaag aagacgttcc aaccacgtct 12900
tcaaagcaag tggattgatg tgacatctcc actgacgtaa gggatgacgc acaatccac 12960
60 tatccttcgc aagacccttc ctctatataa ggaagttcat ttcatttgga gaggacacgc 13020

tcgaggctag catggatctc gggcccaaaa taatgatttt attttgactg atagtgcct 13080
gttcgttgca acaaattgat gagcaatgct tttttataat gccaaacttg tacaaaaaag 13140
5 ctgaacgaga aacgtaaaat gatataaata tcaatatatt aaattagatt ttgcataaaa 13200
aacagactac ataatactgt aaaacacaac atatccagtc actatgaatc aactacttag 13260
atggtattag tgacctgtag tcgaccgaca gccttccaaa tgttcttcgg gtgatgctgc 13320
10 caacttagtc gaccgacagc cttccaaatg ttcttctcaa acggaatcgt cgtatccagc 13380
ctactcgcta ttgtcctcaa tgccgtatta aatcataaaa agaaataaga aaaagagggtg 13440
15 cgagcctctt ttttgtgtga caaaataaaa acatctacct attcatatac gctagtgtca 13500
tagtcctgaa aatcatctgc atcaagaaca atttcacaac tcttatactt ttctcttaca 13560
agtgcgttcgg cttcatctgg attttcagcc tctatactta ctaaacgtga taaagtttct 13620
20 gtaatttcta ctgtatcgac ctgcagactg gctgtgtata agggagcctg acatttatat 13680
tccccagaac atcagggtta tggcgttttt gatgtcattt tcgcggtggc tgagatcagc 13740
25 cacttcttcc ccgataacgg agaccggcac actggccata tcggtggtca tcatgcgcca 13800
gctttcatcc ccgatatgca ccaccgggta aagttcacgg gagactttat ctgacagcag 13860
acgtgcactg gccaggggga tcaccatccg tcgcccgggc gtgtcaataa tatcactctg 13920
30 tacatccaca aacagacgat aacggctctc tcttttatag gtgtaaacct taaactgcat 13980
ttcaccagtc cctgttctcg tcagcaaaag agccgttcat ttcaataaac cgggcgacct 14040
35 cagccatccc ttctgatatt tccgctttcc agcgttcggc acgcagacga cgggcttcat 14100
tctgcatggt tgtgcttacc agaccggaga tattgacatc atatatgcct tgagcaactg 14160
atagctgtcg ctgtcaactg tcaactgtaat acgctgcttc atagcacacc tctttttgac 14220
40 atacttcggg tagtgccgat caacgtctca ttttcgcaa aagttggccc agggcttccc 14280
ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacagggtat 14340
45 ttattcggcg caaagtgcgt cgggtgatgc tgccaactta gtcgactaca ggtcactaat 14400
accatctaag tagttgattc atagtgcctg gatatgttgt gttttacagt attatgtagt 14460
ctgtttttta tgcaaaatct aatttaatat attgatattt atatcatttt acgtttctcg 14520
50 ttcagcttcc ttgtacaaag ttggcattat aagaaagcat tgcttatcaa tttgttgcaa 14580
cgaacaggtc actatcagtc aaaataaaat cattatttgc catccagctg cagctcctcg 14640
55 aggaattcgg taccccaatt ggtaaggaaa taattatttt cttttttcct tttagtataa 14700
aatagttaag tgatgttaat tagtatgatt ataataatat agttgttata attgtgaaaa 14760
aataatttat aaatatattg tttacataaa caacatagta atgtaaaaaa atatgacaag 14820
60 tgatgtgtaa gacgaagaag ataaaagttg agagtaagta tattattttt aatgaatttg 14880

atcgaacatg taagatgata tacggccggt aagaggttcc aactttcacc ataatagaaat 14940
 aagatcacta cgggcggtat tttttgagtt atcgagattt tcaggagcta aggaagctaa 15000
 5 aatggagaaa aaaatcactg gatataccac cgttgatata tcccaatggc atcgtaaaga 15060
 acattttgag gcatttcagt cagttgctca atgtacctat aaccagaccg ttcagctgga 15120
 10 tattacggcc tttttaaaga cgttaaagaa aaataagcac aagttttatc cggcctttat 15180
 tcacattctt gcccgctga tgaatgctca tccggaattc cgtatggcaa tgaaagacgg 15240
 tgagctggtg atatgggata gtgttcaccc ttgttacacc gttttccatg agcaaactga 15300
 15 aacgttttca tcgctctgga gtgaatacca cgacgatttc cggcagtttc tacacatata 15360
 ttcgcaagat gtggcggtgt acggtgaaaa cctggcctat ttccctaaag ggttttattga 15420
 gaatatgttt ttcgtctcag ccaatccctg ggtgagtttc accagttttg atttaaactg 15480
 20 ggccaatatg gacaacttct tcgccccgt tttcaccatg ggcaaatatt atacgcaagg 15540
 cgacaagggtg ctgatgccgc tggcgattca gggtcatcat gccgtctgtg atggcttcca 15600
 25 tgtcggcaga atgcttaatg aattacaaca gtactgcgat gagtggcagg gcggggcgta 15660
 atcgcggtga tccggcttac taaaagccag ataacagtat gcgtatttgc gcgctgattt 15720
 ttgcggtata agaatatata ctgatatgtc gggcccataa tagtaattct agctgggttg 15780
 30 atgaattaaa tatcaatgat aaaatactat agtaaaaata agaataaata aattaaaata 15840
 atattttttt atgattaata gtttattata taattaaata tctataccat tactaaatat 15900
 35 tttagttaa aagttaataa atattttgtt agaaattcca atctgctgt aatttatcaa 15960
 taaacaaaat attaaataac aagctaaagt aacaaataat atcaaactaa tagaacagt 16020
 aatctaagt aacaaaacat aatctaagc taatataaca aagcgcaaga tctatcattt 16080
 40 tatatagtat tattttcaat caacattctt attaatctt aaataatact tgtagtttta 16140
 ttaacttcta aatggattga ctattaatta aatgaattag tcgaacatga ataaacaagg 16200
 45 taacatgata gatcatgtca ttgtgttctc attgatctta catttggtt gattacagtt 16260
 gggaaattgg gttcgaaatc gataagcttg gatcctctag agagctgcag ctggatggca 16320
 aataatgatt ttattttgac tgatagtgc ctgttcgttg caacaaattg ataagcaatg 16380
 50 ctttcttata atgccaaactt tgtacaagaa agctgaacga gaaacgtaaa atgatataaa 16440
 tatcaatata ttaaattaga ttttgcataa aaaacagact acataaact gtaaaacaca 16500
 55 acatatccag tcactatgaa tcaactactt agatggtatt agtgacctgt agtcgactaa 16560
 gttggcagca tcacccgacg cactttgcgc cgaataaata cctgtgacgg aagatcactt 16620
 cgagaataa ataaatcctg gtgtccctgt tgataccggg aagccctggg ccaacttttg 16680
 60 gcgaaaatga gacgttgatc ggcaactacc atttcacaac tcttatactt ttctcttaca 16740

agtcgttcgg cttcatctgg attttcagcc tctatactta ctaaactga taaagtttct 16800
gtaatttcta ctgtatcgac ctgcagactg gctgtgtata agggagcctg acattttatat 16860
5 tccccagaac atcagggttaa tggcggtttt gatgtcattt tcgcggtggc tgagatcagc 16920
cacttcttcc ccgataacgg agaccggcac actggccata tcggtggtca tcatgcgcca 16980
10 gctttcatcc ccgatatgca ccaccgggta aagttcacgg gagactttat ctgacagcag 17040
acgtgcactg gccaggggga tcaccatccg tcgcccgggc gtgtcaataa tatcactctg 17100
tacatccaca aacagacgat aacggctctc tcttttatag gtgtaaacct taaactgcat 17160
15 ttcaccagtc cctgttctcg tcagcaaaag agccgttcat ttcaataaac cgggcgacct 17220
cagccatccc ttcctgattt tccgctttcc agcgttcggc acgcagacga cgggcttcat 17280
tctgcatggt tgtgcttacc agaccggaga tattgacatc atatatgcct tgagcaactg 17340
20 atagctgtcg ctgtcaactg tcaactgtaat acgctgcttc atagcacacc tctttttgac 17400
atacttctgt tcttgatgca gatgattttc aggactatga cactagcgta tatgaatagg 17460
25 tagatgtttt tattttgtca cacaaaaaag aggctcgcac ctctttttct tatttctttt 17520
tatgatttaa tacggcattg aggacaatag cgagtaggct ggatacgacg attccgtttg 17580
agaagaacat ttggaaggct gtcggtcgac taagttggca gcatcaccog aagaacattt 17640
30 ggaaggctgt cggtcgacta caggtcacta ataccatcta agtagttgat tcatagtac 17700
tggatatgtt gtgttttaca gtattatgta gtctgtttt tatgcaaaat ctaatttaat 17760
35 atattgatat ttatatcatt ttacgtttct cggttcagctt ttttgtacaa agttggcatt 17820
ataaaaaagc attgctcatc aatttggtgc aacgaacagg tcactatcag tcaaaaaata 17880
atcattatit ggggcccgag atccatgcta gctctagagt cctgctttta tgagatatgc 17940
40 gagacgccta tgatcgcatg atatttgctt tcaattctgt tgtgcacggt gtaaaaaacc 18000
tgagcatgtg tagctcagat ccttaccgcc ggtttcggtt cattctaata aatatatcac 18060
45 ccgttactat cgtattttta tgaataatat tctccgttca atttactgat tgtaccctac 18120
tacttatatg tacaatatta aaatgaaaac aatatattgt gctgaatagg tttatagcga 18180
catctatgat agagcgccac aataacaaac aattgcgttt tattattaca aatccaattt 18240
50 taaaaaaagc ggcagaaccg gtcaaaccta aaagactgat tacataaatc ttattcaaat 18300
ttcaaaaagg cccaggggct agtatctacg acacaccgag cggcgaaata ataacgttca 18360
55 ctgaagggaa ctccggttcc ccgccggcgc gcatgggtga gattccttga agttgagtat 18420
tggccgtccg ctctaccgaa agttacgggc accattcaac ccggtccagc acggcgggcg 18480
ggtaaccgac ttgctgcccc gagaattatg cagcattttt ttggtgtatg tgggccccaa 18540
60 atgaagtga ggtcaaacct tgacagtac gacaaatcgt tggcggggtc cagggcgaat 18600

tttgcgacaa catgtcgagg ctcagcagga cctgcaggca tgcaagctag cttactagtg 18660
 atgcatattc tatagtgtca cctaaatctg c 18691

5
 <210> 14
 <211> 59
 <212> DNA
 <213> Artificial sequence

10
 <220>
 <223> forward primer used for the amplification of 200 and 400 bp CHS f
 ragments

15 <400> 14
 ggggacaagt ttgtacaaaa aagcaggctg cactgctaac cctgagaacc atgtgcttc 59

20
 <210> 15
 <211> 59
 <212> DNA
 <213> Artificial sequence

25
 <220>
 <223> reverse primer for amplification of 400 bp CHS fragment

<400> 15
 ggggaccact ttgtacaaga aagctgggtc gcttgacgga aggacggaga ccaagaagc 59

30
 <210> 16
 <211> 59
 <212> DNA
 <213> Artificial sequence

35
 <220>
 <223> reverse primer for amplification of 200bp CHS fragment

40
 <400> 16
 ggggaccact ttgtacaaga aagctgggta ggagccatgt aagcacacat gtgtggggtt 59

45
 <210> 17
 <211> 100
 <212> DNA
 <213> Artificial sequence

50
 <220>
 <223> forward primer for amplification of 100bp CHS fragment

<400> 17
 ggggacaagt ttgtacaaaa aagcaggctg cactgctaac cctgagaacc atgtgcttca 60
 ggcggagtat cctgactact acttccgcat caccaacagt 100

55
 <210> 18
 <211> 100
 <212> DNA
 60 <213> Artificial sequence

<220>

<223> reverse primer for amplification of 100 bp CHS fragment

<400> 18
 5 ggggaccact ttgtacaaga aagctgggta acttctcctt gaggtcggtc atgtgttcac 60
 tgttggtgat gcggaagtag tagtcaggat actccgcctg 100

<210> 19
 10 <211> 79
 <212> DNA
 <213> Artificial sequence

<220>
 15 <223> forward primer for amplification of 50 bp CHS fragment

<400> 19
 ggggacaagt ttgtacaaaa aagcaggctg cactgctaac cctgagaacc atgtgcttca 60
 20 ggcggagtat cctgactac 79

<210> 20
 <211> 79
 25 <212> DNA
 <213> Artificial sequence

<220>
 <223> reverse primer for 50 bp CHS fragment

30 <400> 20
 ggggaccact ttgtacaaga aagctgggtg tagtcaggat actccgcctg aagcacatgg 60
 ttctcagggt tagcagtgc 79
 35

<210> 21
 <211> 54
 <212> DNA
 40 <213> Artificial sequence

<220>
 <223> forward primer for amplification of the 25 bp CHS fragment

45 <400> 21
 ggggacaagt ttgtacaaaa aagcaggctg cactgctaac cctgagaacc atgt 54

<210> 22
 50 <211> 54
 <212> DNA
 <213> Artificial sequence

<220>
 55 <223> reverse primer for amplification of the 25 bp CHS fragment

<400> 22
 ggggaccact ttgtacaaga aagctgggta catggttctc agggtagca gtgc 54
 60

<210> 23
 <211> 15

<212> DNA
<213> Artificial sequence

<220>
5 <223> acceptor vector pHELLSGATE4

<400> 23
aaaaaaaaaa aaaaaa 15

10
<210> 24
<211> 17476
<212> DNA
<213> Artificial sequence

15
<220>
<223> acceptor vector pHELLSGATE8

<400> 24

20 ggccgcacta gtgatatccc gcggccatgg cggccgggag catgcgacgt cgggcccaat 60
tcgccctata gtgagtcgta ttacaattca ctggcgcgctg ttttacaacg tcgtgactgg 120
gaaaaccctg gcgttaccca acttaatcgc cttgcagcac atcccccttt cgccagctgg 180
25 cgtaatatgcg aagaggcccg caccgatcgc ccttcccaac agttgcgcag cctgaatggc 240
gaatggaaat tgtaaacggtt aatgggtttc tggagttaa tgagctaagc acatacgtca 300
30 gaaaccatta ttgcgcgttc aaaagtcgcc taaggtcact atcagctagc aaatatttct 360
tgtcaaaaat gctccactga cgttccataa attccctcgc gtatccaatt agagtctcat 420
attcactctc aatccaaata atctgcaatg gcaattacct tatccgcaac ttctttacct 480
35 atttcgcgcc ggatccgggc aggttctccg gccgcttggg tggagaggct attcggctat 540
gactgggcac aacagacaat cggctgctct gatgccgcgc tgttccggct gtcagcgcag 600
40 gggcgcccg tttttttgt caagaccgac ctgtccggtg ccctgaatga actgcaggac 660
gaggcagcgc ggctatcgtg gctggccacg acgggcgttc cttgcgcagc tgtgctcgac 720
gttgctcactg aagcggaag ggactggctg ctattgggcg aagtgcggg gcaggatctc 780
45 ctgtcatctc accttgctcc tgccgagaaa gtatccatca tggctgatgc aatgcggcgg 840
ctgcatacgc ttgatccggc tacctgcca ttcgaccacc aagcgaaaca tcgcatcgag 900
50 cgagcacgta ctcggatgga agccggtctt gtcgatcagg atgatctgga cgaagagcat 960
caggggctcg cgccagccga actgttcgcc aggetcaagg cgcgcagtcg cgacggcgag 1020
gatctcgtcg tgacctatgg cgatgcctgc ttgccgaata tcatggtgga aaatggccgc 1080
55 ttttctggat tcatcgactg tggccggctg ggtgtggcgg accgctatca ggacatagcg 1140
ttggctaccc gtgatattgc tgaagagctt ggcggcgaat gggtgaccg cttcctcgtg 1200
60 ctttacggta tcgccgctcc cgattcgcag cgcacgcct tctatcgcct tcttgacgag 1260
ttcttctgag cgggactctg gggttcgaaa tgaccgacca agcgacgcc aacctgccat 1320

cacgagatTT cgattccacc gccgccttct atgaaaggTT gggcttcgga atcgTTTTcc 1380
5 gggacgccgg ctggatgatc ctccagcgcg gggatctcat gctggagttc ttcgcccacc 1440
ccgatccaac acttacgttt gcaacgtcca agagcaaata gaccacgaac gccggaaggT 1500
tgccgcagcg tgtggattgc gtctcaattc tctcttgagc gaatgcaatg atgaatatga 1560
10 tactgactat gaaactttga ggggaatactg cctagcaccg tcacctcata acgtgcatca 1620
tgcatgccct gacaacatgg aacatcgcta tttttctgaa gaattatgct cgttgaggga 1680
tgtcgccgga attgcagcta ttgccaacat cgaactacc ctcacgcatg cattcatcaa 1740
15 tattattcat gcggggaaag gcaagattaa tccaactggc aaatcatcca gcgtgattgg 1800
taacttcagt tccagcgact tgattcgttt tggTgctacc cacgttttca ataaggacga 1860
20 gatggtggag taaagaagga gtgcgtcgaa gcagatcgTt caaacatttg gcaataaagt 1920
ttcttaagat tgaatcctgt tgccggtctt gcgatgatta tcatataatt tctgttgaat 1980
tacgttaagc atgtaataat taacatgtaa tgcatgacgt tatttatgag atgggtTTTT 2040
25 atgattagag tcccgcaatt atacatttaa tacgcgatag aaaacaaaat atagcgcgca 2100
aactaggata aattatcgcg cgcggtgtca tctatgttac tagatcgaat taattccagg 2160
30 cggTgaaggg caatcagctg ttgcccgtct cactggtgaa aagaaaaacc accccagtac 2220
attaaaaacg tccgcaatgt gttattaagt tgtctaagcg tcaatttgTt tacaccacaa 2280
tatatcctgc caccagccag ccaacagctc cccgaccggc agctcggcac aaaatcacca 2340
35 ctcgatacag gcagcccatc agtccgggac ggcgtcagcg ggagagccgt tgtaaggcgg 2400
cagactttgc tcatgttacc gatgctattc ggaagaacgg caactaagct gccgggtttg 2460
40 aaacacggat gatctcgcg agggtagcat gttgattgta acgatgacag agcgttgctg 2520
cctgtgatca aatatcatct ccctcgaga gatccgaatt atcagccttc ttattcattt 2580
ctcgcttaac cgtgacaggc tgtcgatctt gagaactatg ccgacataat aggaaatcgc 2640
45 tggataaagc cgctgaggaa gctgagtggc gctatttctt tagaagtga cgttgacgat 2700
gtcgacggat cttttccgct gcataaccct gcttcggggT cattatagcg attttttcgg 2760
50 tatatccatc ctttttcgca cgatatacag gattttgcc aagggttcgt gtagactttc 2820
cttggTgtat ccaacggcgt cagccgggca ggataggTga agtaggccca cccgcgagcg 2880
ggTgttcctt cttcactgtc ctttatcgc acctggcggt gctcaacggg aatcctgctc 2940
55 tgcgaggctg gccggctacc gccggcgtaa cagatgaggg caagcggatg gctgatgaaa 3000
ccaagccaac caggggtgat gctgccaaTt tactgattta gtgtatgatg gtgtttttga 3060
60 ggtgctccag tggcttctgt ttctatcagc tgtccctcct gttcagctac tgacggggTg 3120
gtgcgtaacg gcaaaagcac cgccggacat cagcgctatc tctgctctca ctgccgtaaa 3180

acatggcaac tgcagttcac ttacaccgct tctcaaccog gtacgcacca gaaaatcatt 3240
gatatggcca tgaatggcgt tggatgccgg gcaacagccc gcattatggg cgttggcctc 3300
5 aacacgattt tacgtcactt aaaaaactca ggccgcagtc ggtaacctcg cgcatacagc 3360
cgggcagtga cgtcatcgtc tgcgcggaaa tggacgaaca gtggggctat gtcggggcta 3420
10 aatcgcgccca gcgctggctg ttttacgcgt atgacagtct ccggaagacg gttgttgccg 3480
acgtattcgg tgaacgcact atggcgacgc tggggcgctt tatgagcctg ctgtcacctc 3540
ttgacgtggt gatatggatg acggatggct ggccgctgta tgaatcccg cctgaagggaa 3600
15 agctgcacgt aatcagcaag cgatatacgc agcgaattga gcggcataac ctgaatctga 3660
ggcagcacct ggacggcgtg ggacggaagt cgctgtcgtt ctcaaaatcg gtggagctgc 3720
20 atgacaaagt catcgggcat tatctgaaca taaaacacta tcaataagtt ggagtcatta 3780
cccaaccagg aagggcagcc cacctatcaa ggtgtactgc cttccagacg aacgaagagc 3840
gattgaggaa aaggcggcgg cgcccgcat gaggctgtcg gcctacctgc tggccgctcg 3900
25 ccagggtac aaaatcacgg gcgtcgtgga ctatgagcac gtccgcgagc tggcccgcac 3960
caatggcgac ctgggcccgc tgggcccgt gctgaaactc tggctcaccg acgaccgcg 4020
30 cacggcgcg ttcggtgatg ccacgatcct cgccctgtg gcgaagatcg aagagaagca 4080
ggacgagctt ggcaaggcca tgatggcggt ggtccgccc agggcagagc catgactttt 4140
ttagccgcta aaacggccgg ggggtgcgct tgattgcaa gcacgtccc atgcgctcca 4200
35 tcaagaagag cgacttcgct gagctggtat tcgtgcagg caagattcgg aataccaagt 4260
acgagaagga cggccagacg gtctacggga ccgacttcat tgccgataag gtggattatc 4320
40 tggacaccaa ggcaccaggc gggtaaatc aggaataagg gcacattgcc ccggcgtag 4380
tcggggcaat cccgcaagga gggatgaatg atcggacgtt tgaccggaag gcatacaggc 4440
aagaactgat cgacgcgggg ttttccgccc aggatgccga aaccatcgca agccgcaccg 4500
45 tcatgcgtgc gcccgcgaa accttccagt ccgtcggtc gatggtccag caagctacgg 4560
ccaagatcga gcgcgacagc gtgcaactgg ctccccctgc cctgcccgcg ccatcggccg 4620
50 ccgtggagcg ttcgctcgt ctogaacagg aggcggcagg tttggcgaag tcgatgacca 4680
tcgacacgcg aggaactatg acgaccaaga agcgaaaaac cgccggcgag gacctggcaa 4740
aacaggtcag cgaggccaag caggccgctg tgctgaaaca cacgaagcag cagatcaagg 4800
55 aaatgcagct ttccttgttc gatattgcgc cgtggccgga cacgatcgca gcgatgcaa 4860
acgacacggc ccgctctgcc ctgttcacca cgcgcaacaa gaaaatccc cgcgaggcgc 4920
60 tgcaaaaaca ggtcattttc cacgtcaaca aggacgtgaa gatcacctac accggcgctc 4980
agctgcgggc cgacgatgac gaactggtgt ggcagcaggt gttggagtac gcgaagcgca 5040

cccctatcgg cgagccgatc accttcacgt tctacgagct ttgccaggac ctgggctggt 5100
 5 cgatcaatgg ccggtattac acgaaggccg aggaatgcct gtcgcgccta caggcgacgg 5160
 cgatgggctt cacgtccgac cgcgttgggc acctggaatc ggtgtcgctg ctgcaccgct 5220
 tccgcgtcct ggaccgtggc aagaaaacgt cccgttgcca ggtcctgac gacgaggaaa 5280
 10 tcgtcgtgct gtttgctggc gaccactaca cgaaattcat atgggagaag taccgcaagc 5340
 tgtcgccgac ggcccgcagg atgttcgact atttcagctc gcaccgggag ccgtacccgc 5400
 tcaagctgga aaccttccgc ctcatgtgcg gatcggaatc caccgcgctg aagaagtggc 5460
 15 gcgagcaggt cggcgaagcc tgcgaagagt tgcgaggcag cggcctggtg gaacacgcct 5520
 gggccaatga tgacctggtg cattgcaaac gctagggcct tgtggggtca gttccggctg 5580
 20 ggggttcagc agccagcgct ttactggcat ttcaggaaca agcgggcact gctcgacgca 5640
 cttgcttcgc tcagtatcgc tcgggacgca cggcgcgctc tacgaactgc cgataaacag 5700
 aggattaaaa ttgacaattg tgattaaggc tcagattcga cggcttgagg cggccgacgt 5760
 25 gcaggatttc cgcgagatcc gattgtcggc cctgaagaaa gctccagaga tgttcgggtc 5820
 cgtttacgag cacgaggaga aaaagcccat ggaggcgttc gctgaacggt tgcgagatgc 5880
 30 cgtggcattc ggcgcctaca tcgacggcga gatcattggg ctgtcggctt tcaaacagga 5940
 ggacggcccc aaggacgctc acaaggcgca tctgtccggc gttttcgtgg agcccgaaca 6000
 gcgaggccga ggggtcgccg gtatgctgct gcgggcgttg ccggcgggtt tattgtcgt 6060
 35 gatgatcgtc cgacagattc caacgggaat ctgggtgatg cgcattctca tcctcggcgc 6120
 acttaatat tgcctattct ggagcttgtt gtttatttcg gtctaccgcc tgccgggagg 6180
 40 ggtcgcgggc acggtaggag ctgtgcagcc gctgatggc gtgttcatct ctgccgctct 6240
 gctaggtagc ccgatacgat tgatggcggc cctgggggct atttgcgga ctgcgggcgt 6300
 ggcgctgttg gtgttgacac caaacgcagc gctagatcct gtcggcgctc cagcgggcct 6360
 45 ggcgggggag gtttccatgg cgttcggaac cgtgctgacc cgcaagtggc aacctcccgt 6420
 gcctctgctc acctttaccg cctggcaact ggcgccgga ggacttctgc tcgttcagt 6480
 50 agcttttagt tttgatccgc caatcccgat gcctacagga accaatgttc tcggcctggc 6540
 gtggctcggc ctgatcggag cgggtttaac ctacttcctt tggttccggg ggatctcgcg 6600
 actcgaacct acagttgttt ccttactggg ctttctcagc cgggatggcg ctaagaagct 6660
 55 attgccgccg atcttcatat gcggtgtgaa ataccgcaca gatgcgtaag gagaaaatac 6720
 cgcacaggc gctcttcgc ttctcgtc actgactcgc tgcgtcgggt cgttcggctg 6780
 60 cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg tatccacaga atcaggggat 6840
 aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccagggaaccg taaaaaggcc 6900

| | | | | | | | |
|----|------------|------------|-------------|------------|-------------|-------------|------|
| | gcgttgctgg | cgtttttcca | taggctccgc | ccccctgacg | agcatcacia | aaatcgacgc | 6960 |
| | tcaagtcaga | ggtggcgaaa | cccgcacagga | ctataaagat | accaggcggt | ttccccctgga | 7020 |
| 5 | agctccctcg | tgcgctctcc | tggtccgacc | ctgccgctta | ccggatacct | gtccgccttt | 7080 |
| | ctcccttcgg | gaagcgtggc | gctttctcaa | tgctcacgct | gtaggtatct | cagttcgggtg | 7140 |
| 10 | taggtcgttc | gctccaagct | gggctgtgtg | cacgaacccc | ccgttcagcc | cgaccgctgc | 7200 |
| | gccttatccg | gtaactatcg | tcttgagtcc | aaccggtaa | gacacgactt | atcgccactg | 7260 |
| | gcagcagcca | ctggtaacag | gattagcaga | gcgaggtatg | taggcgggtg | tacagagttc | 7320 |
| 15 | ttgaagtgg | ggcctaacta | cggtacact | agaaggacag | tatttggtat | ctgcgctctg | 7380 |
| | ctgaagccag | ttaccttcgg | aaaaagagtt | ggtagctctt | gatccggcaa | acaaaccacc | 7440 |
| 20 | gctggtagcg | gtggtttttt | tggttgcaag | cagcagatta | cgcgagaaa | aaaaggatat | 7500 |
| | caagaagatc | ctttgatctt | ttctacgggg | tctgacgctc | agtggaacga | aaactcacgt | 7560 |
| | taagggatth | tggtcatgag | attatcaaaa | aggatcttca | cctagatcct | tttaaattaa | 7620 |
| 25 | aatgaagtt | ttaaataaat | ctaaagtata | tatgagtaaa | cttggctctga | cagttacca | 7680 |
| | tgcttaata | gtgaggcacc | tatctcagcg | atctgtctat | ttcgttcatc | catagttgcc | 7740 |
| 30 | tgactccccg | tcgtgtagat | aactacgata | cgaggagggt | taccatctgg | ccccagtgtc | 7800 |
| | gcaatgatac | cgcgagaccc | acgtccaccg | gtccagatt | tatcagcaat | aaaccagcca | 7860 |
| | gccggaaggg | ccgagcgag | aagtggctct | gcaactttat | ccgcctccat | ccagtctatt | 7920 |
| 35 | aaacaagtgg | cagcaacgga | ttcgcaaacc | tgctacgcct | tttgtgcaa | aagccgcgcc | 7980 |
| | aggtttcgga | tccgctgtgc | caggcgtag | gcgtcatatg | aagatttcgg | tgatccctga | 8040 |
| 40 | gcaggtggcg | gaaacattgg | atgctgagaa | ccatttcatt | gttcgtgaag | tgttcgatgt | 8100 |
| | gcacctatcc | gaccaaggct | ttgaactatc | taccagaagt | gtgagcccct | accggaagga | 8160 |
| | ttacatctcg | gatgatgact | ctgatgaaga | ctctgcttgc | tatggcgcat | tcatcgacca | 8220 |
| 45 | agagcttgte | gggaagattg | aactcaactc | aacatggaac | gatctagcct | ctatcgaa | 8280 |
| | cattgttggt | tcgcacacgc | accgaggcaa | aggagtcgcg | cacagtctca | tcgaatttgc | 8340 |
| 50 | gaaaaagtgg | gcactaagca | gacagctcct | tggcatacga | ttagagacac | aaacgaacaa | 8400 |
| | tgtacctgcc | tgcaatttgt | acgcaaaatg | tggttttact | ctcggcggca | ttgacctgtt | 8460 |
| | cacgtataaa | actagacctc | aagtctcgaa | cgaaacagcg | atgtactgg | actggttctc | 8520 |
| 55 | gggagcacag | gatgacgcct | aacaattcat | tcaagccgac | accgcttcgc | ggcgcggtt | 8580 |
| | aattcaggag | ttaaacatca | tgagggaagc | ggtgatcgcc | gaagtatcga | ctcaactatc | 8640 |
| 60 | agaggtagtt | ggcgctcatc | agcgccatct | cgaaccgacg | ttgctggccg | tacatttgta | 8700 |
| | cggtccgca | gtggatggcg | gcctgaagcc | acacagtgat | attgatttgc | tggttacgg | 8760 |

gaccgtaagg cttgatgaaa caacgcggcg agctttgatc aacgaccttt tggaaacttc 8820
 5 ggcttcccct ggagagagcg agattctccg cgctgtagaa gtcaccattg ttgtgcacga 8880
 cgacatcatt ccgtggcggtt atccagctaa gcgcgaactg caatttggag aatggcagcg 8940
 caatgacatt cttgcaggtg tcttcgagcc agccacgatc gacattgatc tggctatctt 9000
 10 gctgacaaaa gcaagagaac atagcggttc cttggtaggt ccagcgggcg aggaactctt 9060
 tgatccgggtt cctgaacagg atctatttga ggcgctaaat gaaaccttaa cgctatggaa 9120
 15 ctgcgcggcc gactgggctg gcgatgagcg aaatgtagtg cttacgttgt cccgcatttg 9180
 gtacagcgca gtaaccggca aaatcgcgcc gaaggatgtc gctgccgact gggcaatgga 9240
 gcgcctgcgc gccagtatc agcccgctat acttgaagct aggcaggctt atcttgga 9300
 20 agaagatcgc ttggcctcgc gcgcagatca gttggaagaa tttgttcaact acgtgaaagg 9360
 cgagatcacc aaggtagtcg gcaaataatg tctaacaatt cgttcaagcc gacgccgctt 9420
 cgcggcgcgc cttactcaa gcgttagaga gctggggaag actatgcgcg atctgttgaa 9480
 25 ggtggttcta agcctcgtac ttgcgatggc atcggggcag gcaacttgctg acctgccaat 9540
 tgttttagtg gatgaagctc gtcttcccta tgactactcc ccatccaact acgacatttc 9600
 30 tccaagcaac tacgacaact ccataagcaa ttacgacaat agtccatcaa attacgacaa 9660
 ctctgagagc aactacgata atagttcatc caattacgac aatagtcgca acggaaatcg 9720
 taggcttata tatagcgcaa atgggtctcg cactttcgcc ggctactacg tcattgcaa 9780
 35 caatgggaca acgaacttct tttccacatc tggcaaaagg atgttctaca ccccaaaagg 9840
 ggggcgcggc gtctatggcg gcaaagatgg gagcttctgc ggggcattgg tcgtcataaa 9900
 40 tggccaattt tcgcttgccc tgacagataa cggcctgaag atcatgtatc taagcaacta 9960
 gcctgctctc taataaaatg ttaggagctt ggctgccatt tttggggtga ggccgttcgc 10020
 ggccgagggg cgagccctt ggggggatgg gagggccgcg ttagcgggcc gggaggggtc 10080
 45 gagaaggggg ggcaccccc ttccggctgc gcggtcacgc gccagggcgc agccctggtt 10140
 aaaaacaagg tttataaata ttggtttaa agcagggtta aagacagggt agcgggtggc 10200
 50 gaaaaacggg cggaaccct tgcaaatgct ggattttctg cctgtggaca gccctcaaa 10260
 tgtcaatagg tgcccccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 10320
 ccctcatctg tcagtagtcg cggccctcaa gtgtcaatac cgagggcac ttatccccag 10380
 55 gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgca 10440
 ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 10500
 60 gccgggtgag tcggccctc aagtgtcaac gtccgcccct catctgtcag tgagggccaa 10560
 gttttccgcg aggtatccac aacgcggcg gcggccgcg gtgtctcgca cacggcttcg 10620

acggcggtttc tggcgcggttt gcaggggcat agacggccgc cagcccagcg gcgagggcaa 10680
ccagcccgggt gagcgtcgga aagggtcgac atcttgctgc gttcggatat tttcgtggag 10740
5 tccccgccac agaccggat tgaaggcgag atccagcaac tcgcgccaga tcatcctgtg 10800
acggaactttt ggcgcgtgat gactggccag gacgtcggcc gaaagagcga caagcagatc 10860
10 acgatttttcg acagcgtcgg atttgcgatc gaggattttt cggcgctgcg ctacgtccgc 10920
gaccgcgttg agggatcaag ccacagcagc ccaactcgacc ttctagccga cccagacgag 10980
ccaagggatc tttttggaat gctgctccgt cgtcaggctt tccgacgttt ggggtggtga 11040
15 acagaagtca ttatcgtacg gaatgccagc actcccagg ggaaccctgt ggttgccatg 11100
cacatacaaa tggacgaacg gataaacctt ttacgcctt tttaaatata cgttattcta 11160
20 ataaacgctc ttttctctta ggtttaccgc ccaatatata ctgtcaaaca ctgatagttt 11220
aaactgaagg cgggaaaacga caatctgatc atgagcggag aattaaggga gtcacgttat 11280
gacccccgcc gatgacgcgg gacaagccgt tttacgtttg gaactgacag aaccgcaacg 11340
25 attgaaggag ccaactcagc ccaatacgca aaccgcctct ccccgcgctt tggccgattc 11400
attaatgcag ctggcacgac aggtttcccg actggaaagc gggcagtgag cgcaacgcaa 11460
30 ttaatgtgag ttagctcact cattaggcac cccaggcttt acactttatg cttccggctc 11520
gtatgttggtg tggaattgtg agcggataac aatttcacac aggaaacagc tatgaccatg 11580
attacgcaa gctatttagg tgacactata gaataactca gctatgcac caacgcgttg 11640
35 ggagctctcc catatcgacc tgcaggcggc cgctcgacga attaatcca atcccacaaa 11700
aatctgagct taacagcaca gttgctcctc tcagagcaga atcgggtatt caacaccctc 11760
40 atatcaacta ctacgttggtg tataacggtc cacatgccgg tatatacgat gactgggggtt 11820
gtacaaaggc ggcaacaaac ggcgttcccg gagggtgcaca caagaaattt gccactatta 11880
cagaggcaag agcagcagct gacgcgtaca caacaagtca gcaaacagac aggttgaact 11940
45 tcatccccaaggaggaaagct caactcaagc ccaagagctt tgctaaggcc ctaacaagcc 12000
caccaaagca aaaagccac tggctcacgc taggaacca aaggccagc agtgatccag 12060
50 ccccaaaaga gatctccttt gccccggaga ttacaatgga cgatttcctc tatctttacg 12120
atctaggaag gaagttcgaa ggtgaagggt acgacactat gttcaccact gataatgaga 12180
aggttagcct cttcaatttc agaaagaatg ctgaccaca gatgggttaga gaggcctacg 12240
55 cagcaggtct catcaagacg atctaccgca gtaacaatct ccaggagatc aaataccttc 12300
ccaagaaggt taaagatgca gtcaaaagat tcaggactaa ttgcatcaag aacacagaga 12360
60 aagacatatt tctcaagatc agaagtacta ttccagtatg gacgattcaa ggcttgcttc 12420
ataaaccaag gcaagtaata gagattggag tctctaaaaa ggtagttcct actgaatcta 12480

aggccatgca tggagtctaa gattcaaatac gaggatctaa cagaactcgc cgtgaagact 12540
5 ggcgaacagt tcatacagag tcttttacga ctcaatgaca agaagaaaat cttcgtcaac 12600
atggtggagc acgacactct ggtctactcc aaaaatgtca aagatacagt ctcagaagac 12660
caaagggcta ttgagacttt tcaacaaagg ataatttcgg gaaacctcct cggattccat 12720
10 tgcccagcta tctgtcactt catcgaaagg acagtagaaa aggaagggtg ctcctacaaa 12780
tgccatcatt gcgataaagg aaaggctatc attcaagatc tctctgccga cagtgggtccc 12840
aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc aaccacgtct 12900
15 tcaaagcaag tggattgatg tgacatctcc actgacgtaa gggatgacgc acaatccac 12960
tatccttcgc aagacccttc ctctatataa ggaagttcat ttcatttggga gaggacacgc 13020
20 tcgagacaag tttgtacaaa aaagctgaac gagaaacgta aaatgatata aatatcaata 13080
tattaaatta gattttgcat aaaaaacaga ctacataata ctgtaaaaca caacatatcc 13140
agtcaactatg aatcaactac ttagatggta ttagtgacct gtagtcgacc gacagccttc 13200
25 caaatgttct tcgggtgatg ctgccaaact agtcgaccga cagccttcca aatgttcttc 13260
tcaaacggaa tcgtcgtatc cagcctactc gctattgtcc tcaatgccgt attaaatcat 13320
30 aaaaagaaat aagaaaaaga ggtgcgagcc tcttttttgt gtgacaaaat aaaaacatct 13380
acctattcat atacgctagt gtcatagtcc tgaaaatcat ctgcatcaag aacaatttca 13440
caactcttat acttttctct tacaagtcgt tcggcttcat ctggattttc agcctctata 13500
35 cttactaaac gtgataaagt ttctgtaatt tctactgtat cgacctgcag actggctgtg 13560
tataaggag cctgacattt atattcccca gaacatcagg ttaatggcgt ttttgatgtc 13620
40 attttcgcgg tggctgagat cagccacttc ttccccgata acggagaccg gcacactggc 13680
catatcgggtg gtcacatgc gccagcttc atccccgata tgaccaccg ggtaaagttc 13740
acgggagact ttatctgaca gcagacgtgc actggccagg gggatcacca tccgtcgccc 13800
45 gggcgtgtca ataatacac tctgtacatc cacaacaga cgataacggc tctctctttt 13860
ataggtgtaa accttaaact gcatttcacc agtccctgtt ctcgtcagca aaagagccgt 13920
50 tcatttcaat aaaccgggcg acctcagcca tcccttctctg attttccgct ttccagcgtt 13980
cggcacgcag acgacgggt tcattctgca tggttgtgct taccagaccg gagatattga 14040
catcatatat gccttgagca actgatagct gtcgctgtca actgtcactg taatacgtg 14100
55 cttcatagca cacctctttt tgacatactt cgggtagtgc cgatcaacgt ctcattttcg 14160
ccaaaagttg gcccagggt tcccgtatc aacagggaca ccaggattta tttattctgc 14220
60 gaagtgatct tccgtcacag gtatttatc gccgcaaagt gcgtcgggtg atgctgcaa 14280
cttagtcgac tacaggtcac taataccatc taagtagttg attcatagt actggatatg 14340

ttgtgtttta cagtattatg tagtctgttt tttatgcaaa atctaattta atatattgat 14400
 atttatatca ttttacgttt ctcgttcagc tttcttgtac aaagtgggtct cgaggaattc 14460
 5 ggtaccccag cttggttaagg aaataattat tttctttttt ccttttagta taaaatagtt 14520
 aagtgatgtt aattagtatg attataataa tatagttggt ataattgtga aaaaataatt 14580
 10 tataaatata ttgtttacat aaacaacata gtaatgtaaa aaaatatgac aagtgatgtg 14640
 taagacgaag aagataaaag ttgagagtaa gtatattatt tttaatgaat ttgatcgaac 14700
 atgtaagatg atatactagc attaatattt gttttaatca taatagtaat tctagctggt 14760
 15 ttgatgaatt aaatatcaat gataaaatac tatagtaaaa ataagaataa ataaattaaa 14820
 ataatatttt tttatgatta atagttttatt atataattaa atatctatac cattactaaa 14880
 20 tatttttagtt taaaagttaa taaatatattt gttagaaatt ccaatctgct tgtaatttat 14940
 caataaacia aatatttaaat aacaagctaa agtaacaaat aatatcaaac taatagaaac 15000
 agtaatctaa tgtaacaaaa cataatctaa tgctaataa acaaagcgca agatctatca 15060
 25 ttttatatag tattattttc aatcaacatt cttattaatt tctaaataat actttagttt 15120
 ttattaactt ctaaatggat tgactattaa ttaaatgaat tagtcgaaca tgaataaaca 15180
 30 aggtaacatg atagatcatg tcatttgtgt atcattgatc ttacatttgg attgattaca 15240
 gttgggaagc tgggttcgaa atcgataagc ttggatcctc tagaccactt tgtacaagaa 15300
 agctgaacga gaaacgtaaa atgatataaa tatcaatata ttaaattaga ttttgcataa 15360
 35 aaaacagact acataatact gtaaaacaca acatatccag tcaactatgaa tcaactactt 15420
 agatggtatt agtgacctgt agtcgactaa gttggcagca tcacccgacg cactttgcgc 15480
 40 cgaataaata cctgtgacgg aagatcactt cgcagaataa ataaatcctg gtgtccctgt 15540
 tgataccggg aagccctggg ccaacttttg gcgaaaatga gacgttgatc ggatttcaca 15600
 actcttatac tttctcttta caagtcgttc ggcttcatct ggattttcag cctctatact 15660
 45 tactaaacgt gataaagttt ctgtaatttc tactgtatcg acctgcagac tggctgtgta 15720
 taagggagcc tgacatttat attccccaga acatcagggt aatggcggtt ttgatgtcat 15780
 50 tttcgcggtg gctgagatca gccacttctt ccccgataac ggagaccggc aactggcca 15840
 tatcggtggt catcatgcgc cagctttcat cccgatatg caccaccggg taaagttcac 15900
 gggagacttt atctgacagc agacgtgcac tggccagggg gatcaccatc cgtcgcccg 15960
 55 gcgtgtcaat aatatcactc tgtacatcca caaacagacg ataacggctc tctcttttat 16020
 aggtgtaaac cttaaactgc atttcaccag tccctgttct cgtcagcaaa agagccgttc 16080
 60 atttcaataa accgggagac ctgagccatc ccttctgat tttccgctt ccagcggtcg 16140
 gcacgcagac gacgggcttc attctgcatg gttgtgctta ccagaccgga gatattgaca 16200

tcatatatgc cttgagcaac tgatagctgt cgctgtcaac tgtcactgta atacgctgct 16260
 tcatagcaca cctctttttg acatacttct gttcttgatg cagatgattt tcaggactat 16320
 5 gacactagcg tatatgaata ggtagatggt tttattttgt cacacaaaaa agaggctcgc 16380
 acctcttttt cttatttctt tttatgattt aatacggcat tgaggacaat agcgagtagg 16440
 10 ctggatacga cgattccgtt tgagaagaac atttggaagg ctgtcggtcg actaagttgg 16500
 cagcatcacc cgaagaacat ttggaaggct gtcggtcgac tacaggtcac taataccatc 16560
 taagtagttg attcatagtg actggatagtg ttgtgtttta cagtattatg tagtctgttt 16620
 15 tttatgcaaa atctaattta atatattgat atttatatca ttttacgttt ctcggttcagc 16680
 tttttgtac aaacttgct agagtcctgc tttaatgaga tatgagagac gcctatgac 16740
 20 gcatgatatt tgctttcaat tctgttgctg acgttgtaaa aaacctgagc atgtgtagct 16800
 cagatcctta ccgccggtt cggttcattc taatgaatat atcaccggt actatcgat 16860
 ttttatgaat aatattctcc gttcaattta ctgattgtac cctactactt atatgtacaa 16920
 25 tattaataatg aaaacaatat attgtgctga ataggtttat agcgacatct atgatagagc 16980
 gccacaataa caaacaattg cgttttatta ttacaaatcc aattttaaaa aaagcggcag 17040
 30 aaccgggtcaa acctaaaaga ctgattacat aaatcttatt caaatttcaa aaggccccag 17100
 gggctagtat ctacgacaca ccgagcggcg aactaataac gttcactgaa gggaactccg 17160
 gttccccgcc ggcgcgcatg ggtgagattc cttgaagttg agtattggcc gtccgctcta 17220
 35 ccgaaagtta cgggcaccat tcaaccgggt ccagcacggc ggccgggtaa ccgacttget 17280
 gccccgagaa ttatgcagca tttttttggt gtatgtgggc cccaaatgaa gtgcagggtca 17340
 40 aaccttgaca gtgacgacaa atcgttgggc ggggccaggc cgaattttgc gacaacatgt 17400
 cgaggctcag caggacctgc aggcattgca gctagettac tagtgatgca tattctatag 17460
 45 tgtcacctaa atctgc 17476

<210> 25
 <211> 17458
 <212> DNA
 50 <213> Artificial sequence

<220>
 <223> acceptor vector pHELLSGATE11

55 <400> 25
 ggcgcgacta gtgatatccc gcggccatgg cggccgggag catgacgacgt cgggccaat 60
 tcgccctata gtgagtcgta ttacaattca ctggccgctg ttttacaacg tcgtgactgg 120
 60 gaaaaccctg gcgttaccca acttaatcgc cttgcagcac atcccccttt cgccagctgg 180
 cgtaatagcg aagaggcccg caccgatcgc cttcccaac agttgcgcag cctgaatggc 240

| | | |
|----|--|------|
| | gaatggaaat tgtaaacgtt aatgggtttc tggagttaa tgagctaagc acatacgtca | 300 |
| | gaaaccatta ttgcgcgttc aaaagtcgcc taaggctact atcagctagc aaatatttct | 360 |
| 5 | tgtcaaaaat gctccactga cgttcataa attcccctcg gtatccaatt agagtctcat | 420 |
| | attcactctc aatccaaata atctgcaatg gcaattacct tatccgcaac ttctttacct | 480 |
| 10 | atttcgcgcc ggatccgggc aggttctccg gccgcttggg tggagaggct attcggctat | 540 |
| | gactgggcac aacagacaat cggctgctct gatgccgcg tgttccggct gtcagcgag | 600 |
| | gggcgcccgg ttctttttgt caagaccgac ctgtccgggtg ccctgaatga actgcaggac | 660 |
| 15 | gaggcagcgc ggctatcgtg gctggccacg acgggcgttc cttgcgcagc tgtgctcgac | 720 |
| | gttgtcactg aagcgggaag ggactggctg ctattgggcg aagtgccggg gcaggatctc | 780 |
| 20 | ctgtcatctc accttgctcc tgccgagaaa gtatccatca tggctgatgc aatgcggcgg | 840 |
| | ctgcatacgc ttgatccggc tacctgccca ttcgaccacc aagcgaaca tcgcatcgag | 900 |
| | cgagcacgta ctcgatgga agccggtctt gtcgatcagg atgatctgga cgaagagcat | 960 |
| 25 | caggggctcg cgccagccga actgttcgcc aggtcgaagg cgcgcagcc cgacggcgag | 1020 |
| | gatctcgtcg tgacctatgg cgatgcctgc ttgccgaata tcatggtgga aaatggccgc | 1080 |
| 30 | ttttctggat tcatcgactg tggccggctg ggtgtggcgg accgctatca ggacatagcg | 1140 |
| | ttggctaccc gtgatattgc tgaagagctt ggcggcgaat gggctgaccg cttcctcgtg | 1200 |
| | ctttacggta tcgccgctcc cgattcgcag cgcacgcct tctatgcct tcttgacgag | 1260 |
| 35 | ttcttctgag cgggactctg gggttcgaaa tgaccgacca agcgacgccc aacctgccat | 1320 |
| | cacgagattt cgattccacc gccgccttct atgaaagggt gggcttcgga atcgttttcc | 1380 |
| 40 | gggacgcggc ctggatgacg ctccagcgcg gggatctcat gctggagtgc ttgccccacc | 1440 |
| | ccgatccaac acttacgttt gcaacgtcca agagcaaata gaccacgaac gccggaagg | 1500 |
| | tgccgcagcg tgtggattgc gtctcaattc tctcttgag gaatgcaatg atgaatatga | 1560 |
| 45 | tactgactat gaaactttga gggaatactg cctagcaccg tcacctcata acgtgcatca | 1620 |
| | tgcatgccct gacaacatgg aacatcgcta tttttctgaa gaattatgct cgttgaggga | 1680 |
| 50 | tgtcgcggca attgcagcta ttgccaacat cgaactaccc ctcacgcatg cattcatcaa | 1740 |
| | tattattcat gcggggaaag gcaagattaa tccaactggc aaatcatcca gcgtgattgg | 1800 |
| | taacttcagt tccagcgact tgattcgttt tgggtgctacc caggttttca ataaggacga | 1860 |
| 55 | gatgggtggag taaagaagga gtgcgtcgaa gcagatcggt caaacatttg gcaataaagt | 1920 |
| | ttcttaagat tgaatcctgt tgccggtctt gcgatgatta tcatataatt tctgttgaat | 1980 |
| 60 | tacgttaagc atgtaataat taacatgtaa tgcatgacgt tattttatgag atgggttttt | 2040 |
| | atgattagag tcccgcgaatt atacatttaa tacgcgatag aaaacaaaat atagcgcgca | 2100 |

aactaggata aattatcgcg cgcggtgtca tctatgttac tagatcgaat taattccagg 2160
 5 cgggtgaaggg caatcagctg ttgcccgtct cactggtgaa aagaaaaacc accccagtac 2220
 attaaaaacg tccgcaatgt gttattaagt tgtctaagcg tcaatttggt tacaccacaa 2280
 tatatcctgc caccagccag ccaacagctc cccgaccggc agctcggcac aaaatcacca 2340
 10 ctcgatacag gcagcccatc agtccgggac ggcgtcagcg ggagagccgt tgtaaggcgg 2400
 cagactttgc tcatgttacc gatgctattc ggaagaacgg caactaagct gccgggtttg 2460
 15 aaacacggat gatctcgcg agggtagcat gttgattgta acgatgacag agcgttgctg 2520
 cctgtgatca aatatcatct cctctgcaga gatccgaatt atcagccttc ttattcattt 2580
 ctgccttaac cgtgacagge tgcgatctt gagaactatg ccgacataat aggaaatcgc 2640
 20 tggataaagc cgctgaggaa gctgagtggc gctatttctt tagaagtga cgttgacgat 2700
 gtcgacggat cttttccgct gcataaccct gcttcggggg cattatagcg attttttcgg 2760
 25 tatatccatc ctttttcgca cgatatacag gattttgcc aagggttcgt gtagactttc 2820
 cttggtgtat ccaacggcgt cagccgggca ggataggtga agtaggccca cccgcgagcg 2880
 ggtgttcctt cttcaactgc ccttattcgc acctggcggg gctcaacggg aatcctgctc 2940
 30 tgcgaggctg gccggctacc gccggcgtaa cagatgaggg caagcggatg gctgatgaaa 3000
 ccaagccaac caggggtgat gctgccaaact tactgattta gtgtatgatg gtgtttttga 3060
 35 ggtgctccag tggtttctgt ttctatcagc tgctcctcct gttcagctac tgacgggggtg 3120
 gtgcgtaacg gcaaaagcac cgccggacat cagcgctatc tctgctctca ctgccgtaaa 3180
 acatggcaac tgcagttcac ttacaccgct tctcaaccog gtacgcacca gaaaatcatt 3240
 40 gatatggcca tgaatggcgt tggatgccgg gcaacagccc gcattatggg cgttggcctc 3300
 aacacgattt tacgtcactt aaaaaactca ggccgcagtc ggtaacctcg cgcatacagc 3360
 45 cgggcagtgat cgtcatcgtc tgcgcggaaa tggacgaaca gtggggctat gtcggggcta 3420
 aatcgcgcca gcgctggctg ttttacgcgt atgacagtct ccggaagacg gttgttgccg 3480
 acgtattcgg tgaacgcact atggcgacgc tggggcgtct tatgagcctg ctgtcacctc 3540
 50 ttgacgtggt gatatggatg acggatggct ggccgctgta tgaatccgc ctgaagggaa 3600
 agctgcacgt aatcagcaag cgatatacgc agcgaattga gcggcataac ctgaatctga 3660
 55 ggcagcacct ggcacggctg ggacggaagt cgctgtcgtt ctcaaaatcg gtggagctgc 3720
 atgacaaagt catcgggcat tatctgaaca taaaacacta tcaataagtt ggagtcatta 3780
 cccaaccagg aagggcagcc cacctatcaa ggtgtactgc cttccagacg aacgaagagc 3840
 60 gattgaggaa aagggcggcg cgccggcat gagcctgtcg gcctacctgc tggccgtcg 3900
 ccagggtac aaaatcacgg gcgtcgtgga ctatgagcac gtccgcgagc tggcccgcat 3960

caatggcgac ctgggccgcc tggggcgccct gctgaaactc tggctcaccg acgacccgcg 4020
5 cacggcgcggt ttcggtgatg ccacgacccg cgccctgctg gcgaagatcg aagagaagca 4080
ggacgagcct ggcaaggcca tgatggcggt ggtccgcccg agggcagagc catgactttt 4140
ttagccgcta aaacggcccg ggggtgcgct tgattgcaa gcacgtcccc atgcgtcca 4200
10 tcaagaagag cgacttcgct gagctggtat tcgtgcaggg caagattcgg aataccaagt 4260
acgagaagga cggccagacg gtctacggga ccgacttcat tgccgataag gtggattatc 4320
15 tggacaccaa ggcaccaggc gggtaaatc aggaataagg gcacattgcc ccggcgtgag 4380
tcggggcaat cccgcaagga ggggtgaatga atcggacgtt tgaccggaag gcatacaggc 4440
aagaactgat cgacgcgggg ttttccgccc aggatgccga aaccatcgca agccgcaccg 4500
20 tcatgcgtgc gcccgcgaa accttcacgt ccgtcggctc gatggtccag caagctacgg 4560
ccaagatcga gcgcgacagc gtgcaactgg ctccccctgc cctgcccgcg ccacgcggcg 4620
ccgtggagcg ttcgctcgt ctgcaacagg aggcggcagg tttggcgaag tcgatgacca 4680
25 tcgacacgcg aggaactatg acgaccaaga agcgaaaaac cgccggcgag gacctggcaa 4740
aacaggtcag cgaggccaag caggccgctg tgctgaaaca cacgaagcag cagatcaagg 4800
30 aaatgcagct ttccttgctc gatattgcgc cgtggccgga cagcatgcca gcgatgcaa 4860
acgacacggc ccgctctgcc ctgttcacca cgcgcaaca gaaaatcccg cgcgaggcgc 4920
tgcaaaaaca ggtcattttc cacgtcaaca aggacgtgaa gatcacctac accggcgtcg 4980
35 agctgcgggc cgacgatgac gaactggtgt ggcagcaggt gttggagtac gcgaagcgca 5040
cccctatcgg cgagccgac accttcacgt tctacgagct ttgccaggac ctgggctggt 5100
40 cgatcaatgg ccggtattac acgaaggccg aggaatgcct gtcgcgccta caggcgacgg 5160
cgatgggctt cagctccgac cgcgttgggc acctggaatc ggtgtcgtg ctgcaccgct 5220
tccgcgtcct ggaccgtggc aagaaaacgt cccgttgcca ggtcctgatc gacgaggaaa 5280
45 tcgtcgtgct gtttctggc gaccactaca cgaaattcat atgggagaag taccgaagc 5340
tgtcgcgcgac gggccgacgg atgttcgact atttcagctc gcaccgggag ccgtacccgc 5400
50 tcaagctgga aaccttcgct ctcatgtgct gatcggatc caccgcgtg aagaagtggc 5460
gcgagcaggt cggcgaagcc tgcgaagagt tgcgaggcag cggcctggtg gaacacgcct 5520
gggtcaatga tgacctggtg cattgcaaac gctagggcct tgtggggtca gttccggctg 5580
55 ggggttcagc agccagcgt ttactggcat ttcaggaaca agcgggcact gtcgacgca 5640
cttgcttcgc tcagtatcgc tcgggacgca cggcgcgctc tacgaactgc cgataaacag 5700
60 aggattaaaa ttgacaattg tgattaaggc tcagattcga cggcttgag cggccgacgt 5760
gcaggatttc cgcgagatcc gattgtcggc cctgaagaaa gctccagaga tgttcgggtc 5820

| | | |
|----|---|------|
| | cgtttacgag cacgaggaga aaaagcccat ggaggcggtc gctgaacggt tgcgagatgc | 5880 |
| 5 | cgtaggcattc ggcgcctaca tcgacggcga gatcattggg ctgtcgggtct tcaaacagga | 5940 |
| | ggacggcccc aaggacgctc acaaggcgca tctgtccggc gttttcgtgg agcccgaaca | 6000 |
| | gcgaggccga ggggtcgccg gtatgctgct gcgggcggtg ccggcgggtt tattgctcgt | 6060 |
| 10 | gatgatcgctc cgacagattc caacgggaat ctggtggatg cgcattctca tcctcggcgc | 6120 |
| | acttaatat tgcctattct ggagcttggt gtttatttcg gtctaccgcc tgcggggcgg | 6180 |
| 15 | ggtcgcggcg acggtaggcg ctgtgcagcc gctgatggc gtgttcattct ctgccgtct | 6240 |
| | gctaggtagc ccgatacgat tgatggcggt cctgggggct atttgcgga ctgccggcgt | 6300 |
| | ggcgctggtg gtgttgacac caaacgcagc gctagatcct gtcggcgctc cagcgggcct | 6360 |
| 20 | ggcgggggcg gtttccatgg cggtcggaac cgtgctgacc cgcaagtggc aaacctccgt | 6420 |
| | gcctctgctc acctttaccg cctggcaact ggcgccgga ggacttctgc tcgttcagt | 6480 |
| 25 | agcttttagtg tttgatccgc caatccgat gcctacagga accaatgttc tcggcctggc | 6540 |
| | gtggctcggc ctgatcggag cgggtttaac ctacttcctt tggttccggg ggatctcgcg | 6600 |
| | actcgaacct acagttgttt cttactggg ctttctcagc cgggatggcg ctaagaagct | 6660 |
| 30 | attgccgcgc atcttcatat gcggtgtgaa ataccgcaca gatgcgtaag gagaaaatac | 6720 |
| | cgcatacggc gctcttcgc tcctcgcctc actgactcgc tgcgctcggc cgttcggctg | 6780 |
| 35 | cggcgagcgg tatcagctca ctcaaaggcg gtaatacggc tatccacaga atcaggggat | 6840 |
| | aacgcaggaa agaactatgt agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc | 6900 |
| | gcgttgctgg cgtttttcca taggctccgc cccctgacg agcatcaca aaatcgacgc | 6960 |
| 40 | tcaagtcaga ggtggcgaaa ccgacagga ctataaagat accaggcggt tccccctgga | 7020 |
| | agctccctcg tgcgtctctc tgttcgacc ctgccgtta ccggatacct gtccgccttt | 7080 |
| 45 | ctcccttcgg gaagcgtggc gctttctcaa tgetcacgct gtaggtatct cagttcggtg | 7140 |
| | taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc | 7200 |
| | gccttatccg gtaactatcg tcttgagtc aaccggtaa gacacgactt atcgccactg | 7260 |
| 50 | gcagcagcca ctggtaacag gattagcaga gcgaggatg taggcgggtg tacagagttc | 7320 |
| | ttgaagtggg ggcctaacta cggctacact agaaggacag tatttggtat ctgcgtctg | 7380 |
| 55 | ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc | 7440 |
| | gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgccagaaa aaaaggatat | 7500 |
| | caagaagatc ctttgatctt ttctacggg tctgacgctc agtggaacga aaactcacgt | 7560 |
| 60 | taagggtatt tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa | 7620 |
| | aaatgaagtt ttaaatacat ctaaagtata tatgagtaaa cttggtctga cagttaccaa | 7680 |

tgcttaatca gtgaggcacc tatctcagcg atctgtctat ttcgttcac catagttgcc 7740
 5 tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg cccagtgct 7800
 gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca 7860
 gccggaaggg ccgagcgagc aagtggctct gcaactttat ccgcctccat ccagtctatt 7920
 10 aaacaagtgg cagcaacgga ttgcgaaacc tgtcacgcct tttgtgcca aagccgcgcc 7980
 aggtttgcga tccgctgtgc caggcggttag gcgtcatatg aagatttcgg tgatccctga 8040
 15 gcaggtggcg gaaacattgg atgctgagaa ccatttcatt gttcgtgaag tgttcgatgt 8100
 gcacctatcc gaccaaggct ttgaactatc taccagaagt gtgagcccct accggaagga 8160
 ttacatctcg gatgatgact ctgatgaaga ctctgcttgc tatggcgcat tcatcgacca 8220
 20 agagcttgtc gggaagattg aactcaactc aacatggaac gatctagcct ctatcgaaaca 8280
 cattgttgtg tcgcacacgc accgagggca aggagtcgcg cacagtctca tcgaatttgc 8340
 25 gaaaaagtgg gcaactaagca gacagctcct tggcatacga ttagagacac aaacgaacaa 8400
 tgtacctgcc tgcaatttgt acgcaaaatg tggctttact ctggcgcgca ttgacctgtt 8460
 cacgtataaa actagacctc aagtctcgaa cgaaacagcg atgtactggg actggttctc 8520
 30 gggagcacag gatgacgcct aacaattcat tcaagccgac accgcttcgc ggcgcggtt 8580
 aattcaggag ttaaacatca tgagggaagc ggtgatcgcc gaagtatcga ctcaactatc 8640
 35 agaggtagtt ggcgtcatcg agcgccatct cgaaccgacg ttgctggccg tacatttgta 8700
 cggctccgca gtggatggcg gcctgaagcc acacagtgat attgatttgc tggttacggt 8760
 gaccgtaagg cttgatgaaa caacgcggcg agctttgatc aacgacctt tggaacttc 8820
 40 ggcttcccct ggagagagcg agattctccg cgctgtagaa gtcaccattg ttgtgcacga 8880
 cgacatcatt ccgtggcggt atccagctaa gcgcgaactg caatttggag aatggcagcg 8940
 45 caatgacatt cttgcaggta tcttcgagcc agccacgac gacattgatc tggtatctt 9000
 gctgacaaaa gcaagagaac atagcgttgc cttggtaggt ccagcggcgg aggaactctt 9060
 tgatccggtt cctgaacagg atctatttga ggcgctaaat gaaaccttaa cgctatggaa 9120
 50 ctgcgcgcc gactgggctg gcgatgagcg aaatgtagtg cttacgttgt cccgcatttg 9180
 gtacagcgca gtaaccggca aaatcgcgcc gaaggatgtc gctgccgact gggcaatgga 9240
 gcgcctgcg gccagtatc agcccgctcat acttgaagct aggcaggctt atcttggaca 9300
 55 agaagatcgc ttggcctcgc gcgcagatca gttggaagaa tttgttcact acgtgaaagg 9360
 cgagatcacc aaggtagtcg gcaaataatg tctaacaatt cgttcaagcc gacgccgctt 9420
 60 cgcggcgcg cttactcaa gcgttagaga gctggggaag actatgcgcg atctgttgaa 9480
 ggtggttcta agcctcgtag ttgcgatggc atcggggcag gcacttgctg acctgccaat 9540

tgtttttagtg gatgaagctc gtcttcccta tgactactcc ccatccaact acgacatttc 9600
 5 tccaagcaac tacgacaact ccataagcaa ttacgacaat agtccatcaa attacgacaa 9660
 ctctgagagc aactacgata atagtctatc caattacgac aatagtcgca acggaaatcg 9720
 taggcttata tatagcgcaa atgggtctcg cacttttcgcc ggctactacg tcattgccaa 9780
 10 caatgggaca acgaacttct tttccacatc tggcaaaagg atgttctaca ccccaaaagg 9840
 ggggcgcggc gtctatggcg gcaaagatgg gagcttctgc ggggcattgg tcgtcataaa 9900
 tggccaattt tcgcttgccc tgacagataa cggcctgaag atcatgtatc taagcaacta 9960
 15 gcctgctctc taataaaatg ttaggagctt ggctgccatt tttgggggtga ggccgttcgc 10020
 ggccgagggg cgcagcccct ggggggatgg gagggccgcg ttagcggggc gggaggggtc 10080
 20 gagaaggggg ggcaccccc ttcggcgtgc gcggtcacgc gccagggcgc agccctgggt 10140
 aaaaacaagg ttataaata ttggtttaa agcagggtta aagacagggt agcgggtggc 10200
 gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaa 10260
 25 tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 10320
 ccctcatctg tcagtagtcg cggcccctca gtgtcaatac cgcagggcac ttatccccag 10380
 30 gcttgctcac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgga 10440
 ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgcgc 10500
 gccgggtgag tcggcccctc aagtgtcaac gtccgcccct catctgtcag tgagggccaa 10560
 35 gttttccgcg aggtatccac aacgcgcgcg gccggcgcgc gtgtctcgca cacggcttcg 10620
 acggcgtttc tggcgcgttt gcagggccat agacggccgc cagcccagcg gcgagggcaa 10680
 40 ccagcccgtg gagcgtcgga aagggtcgac atcttctgc gttcggatat tttcgtggag 10740
 ttcccgccac agaccggat tgaaggcgag atccagcaac tcgcgccaga tcatctgtg 10800
 acggaacttt ggcgcgtgat gactggccag gacgtcggcc gaaagagcga caagcagatc 10860
 45 acgattttcg acagcgtcgg atttgcgac gaggattttt cggcgcgtcg ctacgtccgc 10920
 gaccgcgttg agggatcaag ccacagcagc cactcgacc ttctagccga ccagacgag 10980
 50 ccaagggatc tttttggaat gctgctccgt cgtcaggctt tccgacgttt ggggtggtga 11040
 acagaagtca ttatcgtagc gaatgccagc actcccagg ggaaccctgt ggttgcatg 11100
 cacatacaaa tggacgaacg gataaacctt ttcacgccct tttaaatac cgttattcta 11160
 55 ataaacgctc ttttctctta ggtttaccgg ccaatatatc ctgtcaaaca ctgatagttt 11220
 aaactgaagg cgggaaacga caatctgatc atgagcggag aattaaggga gtcacgttat 11280
 60 gacccccgcc gatgacgcgg gacaagccgt tttacgtttg gaactgacag aaccgcaacg 11340
 attgaaggag ccactcagcc ccaatacgca aaccgcctct cccgcgcgt tggccgattc 11400

attaatgcag ctggcacgac aggtttcccg actggaaagc gggcagtgag cgcaacgcaa 11460
5 ttaatgtgag ttagctcact cattaggcac cccaggcttt acactttatg cttccggctc 11520
gtatgttggtg tggaaattgtg agcggataac aatttcacac aggaaacagc tatgaccatg 11580
attacgcaa gctatttagg tgacactata gaatactcaa gctatgcac caacgcgttg 11640
10 ggagctctcc catatcgacc tgcaggcggc cgctcgacga attaatcca atcccacaaa 11700
aatctgagct taacagcaca gttgctcctc tcagagcaga atcgggtatt caacaccctc 11760
atatcaacta ctacgttggtg tataacgggc cacatgccgg tatatacgat gactgggggtt 11820
15 gtacaaaggc ggcaacaaac ggcgttcccg gagttgcaca caagaaattt gccactatta 11880
cagaggcaag agcagcagct gacgcgtaca caacaagtca gcaaacagac aggttgaact 11940
20 tcatcccaa aggagaagct caactcaagc ccaagagctt tgctaaggcc ctaacaagcc 12000
caccaaagca aaaagccac tggctcacgc taggaaccaa aaggccagc agtgatccag 12060
ccccaaaaga gatctccttt gccccggaga ttacaatgga cgatttcctc tatctttacg 12120
25 atctaggaag gaagttcgaa ggtgaagggtg acgacactat gttcaccact gataatgaga 12180
aggttagcct cttcaatttc agaaagaatg ctgaccaca gatggttaga gaggcctacg 12240
30 cagcaggtct catcaagacg atctaccoga gtaacaatct ccaggagatc aaataccttc 12300
ccaagaaggt taaagatgca gtcaaaagat tcaggactaa ttgcatcaag aacacagaga 12360
aagacatatt tctcaagatc agaagtacta ttccagtatg gacgattcaa ggcttgcttc 12420
35 ataaaccaag gcaagtaata gagattggag tctctaaaaa ggtagttcct actgaatcta 12480
aggccatgca tggagtctaa gattcaaatc gaggatctaa cagaactcgc cgtgaagact 12540
40 ggcgaacagt tcatacagag tcttttacga ctcaatgaca agaagaaaat cttcgtcaac 12600
atggtggagc acgacactct ggtctactcc aaaaatgtca aagatacagt ctcagaagac 12660
caaagggcta ttgagacttt tcaacaaagg ataatttcgg gaaacctcct cggattccat 12720
45 tgcccagcta tctgtcactt catcgaaagg acagtagaaa aggaagggtg ctctacaaa 12780
tgccatcatt gcgataaagg aaaggctatc attcaagatc tctctgccga cagtgggtccc 12840
50 aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc aaccacgtct 12900
tcaaagcaag tggattgatg tgacatctcc actgacgtaa gggatgacgc acaatccac 12960
tatccttcgc aagacccttc ctctatataa ggaagttcat ttcatTTgga gaggacacgc 13020
55 tcgagacaag tttgtacaaa aaagctgaac gagaaacgta aaatgatata aatatcaata 13080
tattaaatta gatTTTgcat aaaaaacaga ctacataata ctgtaaaaca caacatatcc 13140
60 agtcactatg aatcaactac ttagatggta ttagtgacct gtagtcgacc gacagccttc 13200
caaatgttct tcgggtgatg ctgccaaatt agtcgaccga cagccttcca aatgttcttc 13260

tcaaacggaa tcgtcgtatc cagcctactc gctattgtcc tcaatgccgt attaaatcat 13320
5 aaaaagaaat aagaaaaaga ggtgcgagcc tcttttttgt gtgacaaaat aaaaacatct 13380
acctattcat atacgctagt gtcatagtcc tgaaaatcat ctgcatcaag aacaatttca 13440
caactcttat acttttctct tacaagtcgt tcggcttcat ctggattttc agcctctata 13500
10 cttactaaac gtgataaagt ttctgtaatt tctactgtat cgacctgcag actggctgtg 13560
tataagggag cctgacattt atattcccca gaacatcagg ttaatggcgt ttttgatgtc 13620
attttcgcgg tggtcgagat cagccacttc ttccccgata acggagaccg gcacactggc 13680
15 catatcgggtg gtcacatgc gccagcttcc atccccgata tgcaccaccg ggtaaagtcc 13740
acgggagact ttatctgaca gcagacgtgc actggccagg gggatcacca tccgtcgccc 13800
20 gggcgtgtca ataatatcac tctgtacatc caaaaacaga cgataacggc tctctctttt 13860
ataggtgtaa accttaaaact gcatttcacc agtccctgtt ctgctcagca aaagagccgt 13920
tcatttcaat aaaccgggcg acctcagcca tcccttctctg attttccgct ttccagcgtt 13980
25 cggcacgcag acgacgggct tcattctgca tggttgtgct taccagaccg gagatattga 14040
catcatatat gccttgagca actgatagct gtcgctgtca actgtcactg taatacgctg 14100
30 cttcatagca cacctctttt tgacataact cgggtagtgc cgatcaacgt ctcatcttcg 14160
ccaaaagttg gcccagggct tcccgggtatc aacagggaca ccaggattta tttattctgc 14220
gaagtgatct tccgtcacag gtatttatcc ggcgcaaagt gcgtcgggtg atgctgcca 14280
35 cttagtcgac tacaggtcac taataaccatc taagtagttg attcatagtg actggatatg 14340
ttgtgtttta cagtattatg tagtctgttt tttatgcaa atctaattta atatattgat 14400
40 atttatatca ttttacgttt ctggttcagc tttcttgtac aaagtggctc cgaggaattc 14460
ggtaccaact gtaaggaaat aattattttc ttttttctt ttagtataaa atagttaagt 14520
gatgttaatt agtatgatta taataatata gttgttataa ttgtgaaaaa ataatttata 14580
45 aatatattgt ttacataaac aacatagtaa tgtaaaaaaa tatgacaagt gatgtgtaag 14640
acgaagaaga taaaagttga gagtaagtat attattttta atgaatttga tcgaacatgt 14700
50 aagatgatat actagcatta atatttgttt taatcataat agtaattcta gctggtttga 14760
tgaattaaat atcaatgata aaatactata gtaaaaaata gaataaataa attaaaaata 14820
tattttttta tgattaatag tttattatat aattaaatat ctataccatt actaaatatt 14880
55 ttagttttaa agttaataaa tattttgtta gaaattccaa tctgcttgta atttatcaat 14940
aaacaaaata ttaataaca agctaaagta acaataata tcaaactaat agaaacagta 15000
60 atctaagtga acaaaacata atctaagtct aatataacaa agcgcaagat ctatcatttt 15060
atatagtatt attttcaatc aacattctta ttaatttcta aataatactt gtagttttat 15120

taacttctaa atggattgac tattaattaa atgaattagt cgaacatgaa taaacaagg 15180
5 aacatgatag atcatgtcat tgtgttatca ttgatcttac atttggattg attacagtta 15240
cttaccttaa gcttggatcc tctagaccac tttgtacaag aaagctgaac gagaaacgta 15300
aaatgatata aatatcaata tattaattaa gatatttgcac aaaaaacaga ctacataata 15360
10 ctgtaaaaca caacatatcc agtcactatg aatcaactac ttagatggta ttagtgacct 15420
gtagtcgact aagttggcag catcaccgca cgcactttgc gccgaataaa tacctgtgac 15480
ggaagatcac ttgcgagaat aaataaatcc tgggtgtccct gttgataccg ggaagccctg 15540
15 ggccaacttt tggcgaaaat gagacgttga tcggatttca caactcttat acttttctct 15600
tacaagtcgt tcggcttcat ctggattttc agcctctata ctactaaac gtgataaagt 15660
20 ttctgtaatt tctactgtat cgacctgcag actggctgtg tataaggag cctgacattt 15720
atattcccca gaacatcagg ttaatggcgt ttttgatgtc attttcgcgg tggctgagat 15780
cagccacttc ttccccgata acggagaccg gcacactggc catatcggtg gtcacatgc 15840
25 gccagctttc atccccgata tgcaccaccg ggtaaagtcc acgggagact ttatctgaca 15900
gcagacgtgc actggccagg gggatcacca tccgtcgccc gggcgtgtca ataatacac 15960
30 tctgtacatc cacaacaga cgataacggc tctctctttt ataggtgtaa accttaact 16020
gcatttcacc agtcctgtt ctgctcagca aaagagccgt tcatttcaat aaaccgggct 16080
acctcagcca tcccttcctg attttccgct ttccagcgtt cggcacgcag acgacgggct 16140
35 tcattctgca tggttgtgct taccagaccg gagatattga catcatatat gccttgagca 16200
actgatagct gtcgtgtca actgtcactg taatacgtg cttcatagca cacctctttt 16260
40 tgacatactt ctgttcttga tgcagatgat ttccaggact atgacactag cgtatatgaa 16320
taggtagatg tttttatttt gtcacacaaa aaagaggctc gcacctcttt ttcttatttc 16380
tttttatgat ttaatacggc attgaggaca atagcgagta ggctggatac gacgattccg 16440
45 tttgagaaga acatttgaa ggctgtcggc cgactaagtt ggcagcatca cccgaagaac 16500
atttggaagg ctgtcggctc actacaggtc actaatacca tctaagtagt tgattcatag 16560
50 tgactggata tgttgtgttt tacagtatta tgtagtctgt tttttatgca aaatctaatt 16620
taatatattg atatttatac ctttttacgt ttctcgttca gcttttttgt acaaacttgt 16680
ctagagtcct gctttaatga gatatgcgag acgcctatga tcgcatgata tttgctttca 16740
55 attctgttgt gcacgttgta aaaaacctga gcatgtgtag ctgagatcct taccgccggt 16800
ttcggttcat tctaataaat atatcaccgg ttactatcgt atttttatga ataatttct 16860
60 ccgttcaatt tactgattgt accctactac ttatatgtac aatattaaaa tgaaaacaat 16920
atattgtgct gaataggttt atagcgacat ctatgataga gcgccacaat aacaaacaat 16980

5 tgcgttttat tattacaaat ccaattttaa aaaaagcggc agaaccggc aaacctaata 17040
 gactgattac ataaatctta ttcaaatttc aaaaggcccc aggggctagt atctacgaca 17100
 caccgagcgg cgaactaata acgttcactg aagggaactc cggttccccg cgggcgcgca 17160
 tgggtgagat tccttgaagt tgagtattgg ccgtccgctc taccgaaagt tacgggcacc 17220
 10 attcaacccg gtccagcagc gcggccgggt aaccgacttg ctgccccgag aattatgcag 17280
 catttttttg gtgtatgtgg gcccctaatg aagtgcaggc caaaccttga cagtgcagac 17340
 aaatcggttg gcgggtccag ggcgaatttt gcgacaacat gtcgaggctc agcaggacct 17400
 15 gcaggcatgc aagctagctt actagtgatg catattctat agtgtcacct aaatctgc 17458

 20 <210> 26
 <211> 17681
 <212> DNA
 <213> Artificial sequence

 25 <220>
 <223> acceptor vector pHELLSGATE12

 <400> 26
 ggccgcacta gtgatatccc gcggccatgg cggccgggag catgcgacgt cgggccaat 60
 30 tcgccctata gtgagtcgta ttacaattca ctggccgctc ttttacaacg tcgtgactgg 120
 gaaaaccctg gcgttaccca acttaatcgc cttgcagcac atcccccttt cgccagctgg 180
 cgtaatatgc aagaggcccg caccgatcgc cttcccaac agttgcgcag cctgaatggc 240
 35 gaatggaaat tgtaaacgtt aatgggtttc tggagttaa tgagctaagc acatacgtca 300
 gaaaccatta ttgcgcgttc aaaagtcgcc taaggctact atcagctagc aaatatttct 360
 40 tgtcaaaaat gctccactga cgttcataa attcccctcg gtatccaatt agagtctcat 420
 attcactctc aatccaaata atctgcaatg gcaattacct tatccgcaac ttctttacct 480
 atttcgccc ggatccgggc aggttctccg gccgcttggg tggagaggct attcggctat 540
 45 gactgggcac aacagacaat cggctgctct gatgcgcgcg tgttccggct gtcagcgcag 600
 gggcgcccgg ttctttttgt caagaccgac ctgtccggtg ccctgaatga actgcaggac 660
 50 gaggcagcgc ggctatcgtg gctggccacg acgggcgttc cttgcgcagc tgtgctcgac 720
 gttgctactg aagcgggaag ggactggctg ctattgggag aagtgcggg gcaggatctc 780
 ctgtcatctc accttgctcc tgccgagaaa gtatccatca tggtgatgc aatgcggcgg 840
 55 ctgcatacgc ttgatccggc tacctgccca ttcgaccacc aagcgaaaca tcgcatcgag 900
 cgagcacgta ctcgatgga agccggtctt gtcgatcagg atgatctgga cgaagagcat 960
 60 caggggctcg cgccagccga actgttcgcc aggtcaagg cgcgcagccc cgacggcgag 1020
 gatctcgtcg tgaccatgg cgatgcctgc ttgccgaata tcatggtgga aaatggccgc 1080

ttttctggat tcatcgactg tggccggctg ggtgtggcgg accgctatca ggacatagcg 1140
 ttggctaccc gtgatattgc tgaagagctt ggcggcgaat gggctgaccg cttcctcgtg 1200
 5 ctttacggta tcgccgctcc cgattcgag cgcacgcct tctatcgctt tcttgacgag 1260
 ttcttctgag cgggactctg gggttcgaaa tgaccgacca agcgacgcc aacctgccat 1320
 10 cacgagattt cgattccacc gccgccttct atgaaagggt gggcttcgga atcgttttcc 1380
 gggacgccgg ctggatgac ctcacgcgg gggatctcat gctggagtgc ttcgcccacc 1440
 ccgatccaac acttacgttt gcaacgtcca agagcaaata gaccacgaac gccggaagggt 1500
 15 tgccgcagcg tgtggattgc gtctcaattc tctcttgag gaatgcaatg atgaatatga 1560
 tactgactat gaaactttga gggaaactg cctagcaccg tcacctcata acgtgcatca 1620
 20 tgcagccct gacaacatgg aacatcgcta ttttctgaa gaattatgct cgttgaggga 1680
 tgtcgcggca attgcagcta ttgccaacat cgaactacce ctcacgcatg cattcatcaa 1740
 tattattcat gcggggaaag gcaagattaa tccaactggc aaatcatcca gcgtgattgg 1800
 25 taacttcagt tccagcgact tgattcgttt tgggtgctacc cacgttttca ataaggacga 1860
 gatggtggag taaagaagga gtgcgtcgaa gcagatcgtt caaacatttg gcaataaagt 1920
 30 ttcttaagat tgaatcctgt tgccggtcct gcgatgatta tcatataatt tctgttgaat 1980
 tacgttaagc atgtaataat taacatgtaa tgcagacgt tatttatgag atgggttttt 2040
 atgattagag tcccgcatt atacatttaa tacgcgatag aaaacaaaat atagcgcgca 2100
 35 aactaggata aattatcgcg cgcggtgtca tctatgttac tagatcgaat taattccagg 2160
 cgggtgaagg caatcagctg ttgccgtct cactggtgaa aagaaaaacc accccagtac 2220
 40 attaaaaacg tccgcaatgt gttattaagt tgtctaagcg tcaatttggt tacaccacaa 2280
 tatatcctgc caccagccag ccaacagctc cccgaccggc agctcggcac aaaatcacca 2340
 ctcgatacag gcagcccatc agtccgggac ggcgtcagcg ggagagccgt tgtaaggcgg 2400
 45 cagactttgc tcatgttacc gatgctattc ggaagaacgg caactaagct gccgggtttg 2460
 aaacacggat gatctcgcg agggtagcat gttgattgta acgatgacag agcgttgctg 2520
 50 cctgtgatca aatatcatct ccctcgcaga gatccgaatt atcagccttc ttattcattt 2580
 ctgccttaac cgtgacaggc tgcgatctt gagaactatg ccgacataat aggaaatcgc 2640
 tggataaagc cgctgaggaa gctgagtggc gctatttctt tagaagtga cgttgacgat 2700
 55 gtcgacggat cttttccgct gcataaccct gcttcggggt cattatagcg attttttcgg 2760
 tataatccatc ctttttcgca cgatatacag gattttgcca aagggttcgt gtagactttc 2820
 60 cttggtgtat ccaacggcgt cagccgggca ggatagggtga agtaggccca cccgcgagcg 2880
 ggtgttcctt cttcactgtc ccttattcgc acctggcggt gctcaacggg aatcctgctc 2940

tgcgaggctg gccggtacc gccggcgtaa cagatgaggg caagcggatg gctgatgaaa 3000
 5 ccaagccaac caggggtgat gctgccaaact tactgattta gtgtatgatg gtgtttttga 3060
 ggtgctccag tggcttctgt ttctatcagc tgccctcct gtccagctac tgacggggtg 3120
 gtgcgtaacg gcaaaagcac cgccggacat cagcgtatc tctgctctca ctgccgtaaa 3180
 10 acatggcaac tgcaattcac ttacaccgct tctcaaccgg gtacgcacca gaaaatcatt 3240
 gatattggcca tgaatggcgt tggatgccgg gcaacagccc gcattatggg cgttggcctc 3300
 aacacgattt tacgtcactt aaaaaactca ggccgcagtc ggtaacctcg cgcatacagc 3360
 15 cgggcagtgga cgtcatcgtc tgcgcggaaa tggacgaaca gtggggctat gtcggggcta 3420
 aatcgcgcca gcgtggcgt ttttacgct atgacagtct ccggaagacg gttgttgccg 3480
 20 acgtattcgg tgaacgcact atggcgacgc tggggcgtct tatgagcctg ctgtcacctc 3540
 ttgacgtggt gatattgatg acggatggct ggccgctgta tgaatccgc ctgaaggaa 3600
 agctgcacgt aatcagcaag cgatatacgc agcgaattga gcggcataac ctgaatctga 3660
 25 ggcagcacct ggcacggctg ggacggaagt cgctgtcgtt ctcaaatcg gtggagctgc 3720
 atgacaaagt catcgggcat tatctgaaca taaaacacta tcaataagtt ggagtcatta 3780
 30 cccaaccagg aagggcagcc cacctatcaa ggtgtactgc cttccagacg aacgaagagc 3840
 gattgaggaa aaggcggcgg cgccggcat gagcctgtcg gcctacctgc tggccgtcgg 3900
 ccagggtac aaatcacgg gcgtcgtgga ctatgagcac gtccgcgagc tggcccgcat 3960
 35 caatggcgac ctgggcccgc tgggggcct gctgaaactc tggctcaacg acgaccgcg 4020
 cacggcgcg ttcggtgatg ccacgatcct cgccctgctg gcgaagatcg aagagaagca 4080
 40 ggacgagctt ggcaaggtca tgatggcgt ggtccgcccg agggcagagc catgactttt 4140
 ttagccgcta aaacggccgg ggggtgcgcg tgattgcaa gcacgtccc atgcgtcca 4200
 tcaagaagag cgacttcgcg gagctggtat tcgtgcagg caagattcgg aataccaagt 4260
 45 acgagaagga cggccagacg gtctacggga ccgacttcat tgccgataag gtggattatc 4320
 tggacaccaa ggcaccaggc ggtcaaactc aggaataagg gcacattgcc ccggcgtgag 4380
 50 tcggggcaat cccgcaagga ggtgaaatga atcggacgtt tgaccggaag gcatacaggc 4440
 aagaactgat cgacgcgggg ttttccgccg aggatgccga aaccatcgca agccgcaccg 4500
 tcatgcgtgc gcccgcgaa accttccagt ccgtcggctc gatggtccag caagctacgg 4560
 55 ccaagatcga gcgcgacagc gtgcaactgg ctccccctgc cctgcccgcg ccacggccg 4620
 ccgtggagcg ttcgcgtcgt ctgcaacagg aggcggcagg tttggcgaag tcgatgacca 4680
 60 tcgacacgcg aggaactatg acgaccaaga agcgaaaaac cgccggcgag gacctggcaa 4740
 aacaggctcag cgaggccaag caggccgcgt tgctgaaaca cacgaagcag cagatcaagg 4800

aaatgcagct ttccttggtc gatattgctc cgtggccgga cacgatgcga gcgatgccaa 4860
 5 acgacacggc ccgctctgcc ctgttcacca cgcgcaacaa gaaaatcccg cgcgaggcgc 4920
 tgcaaaacaa ggtcattttc cacgtcaaca aggacgtgaa gatcacctac accggcgctc 4980
 agctgcgggc cgacgatgac gaactggtgt ggcagcaggt gttggagtac gcgaagcgca 5040
 10 cccctatcgg cgagccgac accttcacgt tctacgagct ttgccaggac ctgggctggt 5100
 cgatcaatgg ccggtattac acgaaggccg aggaatgcct gtcgcgccca caggcgacgg 5160
 cgatgggctt cacgtccgac cgcgttgggc acctggaatc ggtgtcgtc ctgcaccgct 5220
 15 tccgcgtcct ggaccgtggc aagaaaacgt cccgttgcca ggtcctgatc gacgaggaaa 5280
 tcgtcgtgct gtttgcgtgc gaccactaca cgaaattcat atgggagaag taccgaagc 5340
 20 tgctgcggac ggcccgacgg atgttcgact atttcagctc gcaccgggag ccgtaccgc 5400
 tcaagctgga aaccttcgc ctcagtgtgc gatcggattc caccgcgtg aagaagtggc 5460
 gcgagcaggt cggcgaagcc tgcgaagagt tgcgaggcag cggcctggtg gaacacgcct 5520
 25 ggggtcaatga tgacctggtg cattgcaaac gctagggcct tgtgggttca gttccggctg 5580
 ggggttcagc agccagcgtt ttactggcat ttcaggaaca agcgggcact gctcgacgca 5640
 30 cttgcttcgc tcagtatcgc tcgggacgca cggcgcgtc tacgaactgc cgataaacag 5700
 aggattaaaa ttgacaattg tgattaaggc tcagattcga cggcttgag cggccgacgt 5760
 gcaggatttc cgcgagatcc gattgtcggc cctgaagaaa gctccagaga tgttcgggtc 5820
 35 cgtttacgag cacgaggaga aaaagcccat ggaggcgtc gctgaacggt tgcgagatgc 5880
 cgtggcattc ggcgcctaca tcgacggcga gatcattggg ctgtcgttct tcaaacagga 5940
 40 ggacggcccc aaggacgctc acaaggcgca tctgtccggc gttttcgtgg agcccgaaca 6000
 gcgaggccga ggggtcgccg gtatgctgct gcgggcgtt cggcggtt tattgctcgt 6060
 gatgatcgtc cgacagattc caacgggaat ctggtggatg cgcattctca tcctcggcgc 6120
 45 acttaataatt tcgctattct ggagcttgtt gtttatttcg gtctaccgcc tgccgggcgg 6180
 ggtcgcggcg acggtaggcg ctgtgcagcc gctgatggc gtgttcattct ctgccgtct 6240
 50 gctaggtagc ccgatacgat tgatggcggc cctgggggct atttgcgaa ctgcgggcgt 6300
 ggcgctgttg gtgttgacac caaacgcagc gctagatcct gtcggcgtc cagcgggcct 6360
 ggcggggggc gtttccatgg cgttcggaac cgtgtgacc cgcaagtggc aacctcccg 6420
 55 gcctctgctc acctttaccg cctggcaact ggcggccgga ggacttctgc tcgttccagt 6480
 agctttagtg tttgatccgc caatcccgat gcctacagga accaatgttc tcggcctggc 6540
 60 gtggctcggc ctgatcggag cgggtttaac ctacttctt tggttccggg ggatctcgg 6600
 actcgaacct acagttgttt ccttactggg ctttctcagc cgggatggc ctaagaagct 6660

attgccgccg atcttcatat gcggtgtgaa ataccgcaca gatgcgtaag gagaaaatac 6720
5 cgcacagagg gctcttccgc ttctcgcgc actgactcgc tgcgctcggc cgttcggctg 6780
cggcgagcgg tatcagctca ctcaaaggcg gtaatacggc tatccacaga atcaggggat 6840
aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc 6900
10 gcgttgctgg cgtttttcca taggctccgc cccctgacg agcatcaca aaatcgacgc 6960
tcaagtcaga ggtggcgaaa ccgacagga ctataaagat accaggcgtt tccccctgga 7020
agctccctcg tgcgctctcc tgttccgacc ctgcgcctta ccggatacct gtccgccttt 7080
15 ctcccttcgg gaagcgtggc gcttttctca tgctcacgct gtaggtatct cagttcggcg 7140
taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc 7200
20 gccttatccg gtaactatcg tcttgagtc aaccggtaa gacacgactt atcgccactg 7260
gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcgggtg tacagagttc 7320
ttgaagtggg ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg 7380
25 ctgaagccag ttaccttcgg aaaaagagtt gtagctctt gatccggcaa acaaaccacc 7440
gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatat 7500
30 caagaagatc ctttgatctt ttctacggg tctgacgctc agtggaacga aaactcacgt 7560
taagggattt tggatcatg attatcaaaa aggatcttca cctagatcct tttaaattaa 7620
aatgaagtt ttaaatcaat ctaaagtata tatgagtaa cttggtctga cagttaccaa 7680
35 tgcttaatca gtgaggcacc tatctcagcg atctgtctat ttcgttcac catagttgcc 7740
tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg cccagtgct 7800
40 gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca 7860
gccggaaggc ccgagcgcag aagtggctct gcaactttat ccgcctccat ccagtctatt 7920
aaacaagtgg cagcaacgga ttcgcaaac tgtcacgcct tttgtgcaa aagccgcgcc 7980
45 aggtttgcga tccgctgtgc caggcgtag gcgtcatatg aagatttcgg tgatccctga 8040
gcaggtggcg gaaacattgg atgctgagaa ccatttcatt gttcgtgaag tgttcgatgt 8100
50 gcacctatcc gaccaaggct ttgaactatc taccagaagt gtgagcccct accggaagga 8160
ttacatctcg gatgatgact ctgatgaaga ctctgcttgc tatggcgcat tcatcgacca 8220
agagcttgtc ggaagattg aactcaactc aacatggaac gatctagcct ctatcgaaca 8280
55 cattgttggt tcgcacacgc accgaggcaa aggagtcgcg cacagtctca tcgaatttgc 8340
gaaaaagtgg gactaagca gacagctcct tggcatacga ttagagacac aaacgaacaa 8400
60 tgtacctgcc tgcaatttgt acgcaaatg tggctttact ctcgcgcgca ttgacctgtt 8460
cacgtataaa actagacctc aagtctcgaa cgaaacagcg atgtactggt actggttctc 8520

gggagcacag gatgacgcct aacaattcat tcaagccgac accgcttcgc ggcgcggtt 8580
aattcaggag ttaaacaatca tgagggaagc ggtgatcgcc gaagtatcga ctcaactatc 8640
5 agaggtagtt ggcgatcatc agcgccatct cgaaccgacg ttgctggcgc tacatttgta 8700
cggtccgca gtggatggcg gcctgaagcc acacagtgat attgatttgc tggttacggt 8760
10 gaccgtaagg cttgatgaaa caacgcggcg agctttgatc aacgacctt tggaaacttc 8820
ggcttccct ggagagagcg agattctccg cgctgtagaa gtcaccattg ttgtgcacga 8880
cgacatcatt ccgtggcggt atccagctaa gcgcgaactg caatttggag aatggcagcg 8940
15 caatgacatt cttgcaggta tcttcgagcc agccacgatc gacattgatc tggctatctt 9000
gctgacaaaa gcaagagaac atagcgttgc cttggtaggt ccagcggcgg aggaactctt 9060
20 tgatccggtt cctgaacagg atctatttga ggcgctaaat gaaaccttaa cgctatggaa 9120
ctcgccgccc gactgggctg gcgatgagcg aaatgtagtg cttacgttgt cccgcatttg 9180
gtacagcgca gtaaccggca aaatcgcgcc gaaggatgtc gctgccgact gggcaatgga 9240
25 gcgcctgccg gccagtatc agcccgcat acttgaagct aggcaggctt atcttgaca 9300
agaagatcgc ttggcctcgc gcgcagatca gttggaagaa tttgttcaact acgtgaaagg 9360
30 cgagatcacc aaggtagtcg gcaaataatg tctaacaatt cgttcaagcc gacgccgctt 9420
cgcgccgcg cttactcaa gcgttagaga gctggggaag actatgcgcg atctgttgaa 9480
ggtggttcta agcctcgtac ttgcgatggc atcggggcag gcacttgctg acctgccaat 9540
35 tgttttagtg gatgaagctc gtcttcccta tgactactcc ccatccaact acgacatttc 9600
tccaagcaac tacgacaact ccataagcaa ttacgacaat agtccatcaa attacgaaa 9660
40 ctctgagagc aactacgata atagttcatc caattacgac aatagtcgca acggaaatcg 9720
taggcttata tatagcgcaa atgggtctcg cactttcgcc ggctactacg tcattgccaa 9780
caatgggaca acgaacttct tttccacatc tggcaaaagg atgttctaca ccccaaaagg 9840
45 ggggcgcggc gtctatggcg gcaaagatgg gagcttctgc ggggcattgg tcgtcataaa 9900
tggccaattt tcgcttgccc tgacagataa cggcctgaag atcatgtatc taagcaacta 9960
50 gcctgctctc taataaaatg ttaggagctt ggctgccatt tttggggtga ggccgttcgc 10020
ggccgagggg cgcagccctt ggggggatgg gagggccgcg ttagcgggcc gggagggttc 10080
gagaaggggg ggcaccccc ttccgctgc gcggtcacgc gccagggcgc agccctggtt 10140
55 aaaaacaagg ttataaata ttggtttaa agcaggtaa aagacaggtt agcgggtggc 10200
gaaaaacggg cggaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaa 10260
60 tgtcaatagg tgccccctc atctgtcagc actctgccc tcaagtgtca aggatcgcgc 10320
ccctcatctg tcagtagtcg cggccctcaa gtgtcaatac cgcagggcac ttatccccag 10380

gcttgtccac atcatctgtg ggaaactcgc gtaaatcag gcgttttcgc cgatttgca 10440
5 ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgcgc 10500
gccgggtgag tcggccctc aagtgtcaac gtccgccct catctgtcag tgaggccaa 10560
gttttccgcg aggtatccac aacgccggcg gccggccgcg gtgtctcgca cacggcttcg 10620
10 acggcgtttc tggcggttt gcagggccat agacggccgc cagcccagcg gcgagggcaa 10680
ccagcccgtt gagcgtcgga aagggtcgac atcttgctgc gttcggatat ttctgtggag 10740
15 ttcccgcac agaccggat tgaaggcgag atccagcaac tcgcgccaga tcatcctgtg 10800
acggaacttt ggcgctgat gactggccag gacgtcgcc gaaagagcga caagcagatc 10860
acgattttcg acagcgtcgg atttgcgac gaggattttt cggcgctgcg ctacgtccgc 10920
20 gaccgcgttg agggatcaag ccacagcagc cactcgacc ttctagccga cccagacgag 10980
ccaagggatc tttttggaat gctgctcgt cgtcaggctt tccgacgtt ggggtggtga 11040
acagaagtca ttatcgtacg gaatgccagc actcccagg ggaaccctgt ggttgcatg 11100
25 cacatacaaa tggacgaacg gataaacctt ttcacgccct tttaaatac cgttattcta 11160
ataaacgctc ttttcttta ggtttaccg ccaatatatc ctgtcaaaca ctgatagttt 11220
30 aaactgaagg cgggaaacga caatctgac atgagcggag aattaaggga gtcacgttat 11280
gacccccgcc gatgacgcg gacaagccgt tttacgtttg gaactgacag aaccgcaacg 11340
attgaaggag cactcagcc ccaatagca aaccgcctct cccgcgcgt tggccgattc 11400
35 attaattgag ctggcagcag aggtttcccg actggaaagc gggcagtga cgcaacgcaa 11460
ttaatgtgag ttagctcact cattaggcac cccaggcttt acactttatg cttccggctc 11520
40 gtatgttggtg tggaattgtg agcggataac aatttcacac aggaaacagc tatgaccatg 11580
attacgcaa gctatttagg tgacactata gaatactcaa gctatgcac caacgcgttg 11640
ggagctctcc catatcgacc tgcaggcggc cgctcgacga attaatcca atcccacaaa 11700
45 aatctgagct taacagcaca gttgctcctc tcagagcaga atcgggtatt caacaccctc 11760
atatcaacta ctacgttggt tataacggtc cacatgccg tatatacgat gactggggtt 11820
50 gtacaaaggc ggcaacaaac ggcgttcccg gagttgcaca caagaaattt gccactatta 11880
cagaggcaag agcagcagct gacgcgtaca caacaagtca gcaaacagac aggttgaact 11940
tcaccccaa aggagaagct caactcaagc ccaagagctt tgctaaggcc ctaacaagcc 12000
55 caccaaagca aaaagccac tggctcacgc taggaaccaa aaggccagc agtgatccag 12060
ccccaaaaga gatctccttt gccccggaga ttacaatgga cgatttctc tatctttacg 12120
60 atctaggaag gaagttcgaa ggtgaagggt acgacactat gttcaccact gataatgaga 12180
aggttagcct cttcaatttc agaaagaatg ctgaccacaca gatgggttaga gaggcctacg 12240

cagcagggtct catcaagacg atctacccga gtaacaatct ccaggagatc aaatcaccttc 12300
ccaagaaggt taaagatgca gtcaaaagat tcaggactaa ttgcatcaag aacacagaga 12360
5 aagacatatt tctcaagatc agaagtacta ttccagtatg gacgattcaa ggcttgcttc 12420
ataaaccaag gcaagtaata gagattggag tctctaaaaa ggtagttcct actgaatcta 12480
10 aggccatgca tggagtctaa gattcaaate gaggatctaa cagaactcgc cgtgaagact 12540
ggcgaacagt tcatacagag tcttttacga ctcaatgaca agaagaaaat cttcgtcaac 12600
atggtggagc acgacactct ggtctactcc aaaaatgtca aagatacagt ctcagaagac 12660
15 caaagggcta ttgagacttt tcaacaaagg ataatttcgg gaaacctcct cggattccat 12720
tgcccagcta tctgtcactt catcgaaagg acagtagaaa aggaaggtgg ctcctacaaa 12780
20 tgccatcatt gcgataaagg aaaggctatc attcaagatc tctctgccga cagtgggtccc 12840
aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc aaccacgtct 12900
tcaaaagcaag tggattgatg tgacatctcc actgacgtaa gggatgacgc acaatcccac 12960
25 tatccttcgc aagacccttc ctctatataa ggaagttcat ttcatttga gaggacacgc 13020
tcgagacaag tttgtacaaa aaagctgaac gagaaacgta aaatgatata aatatcaata 13080
30 tattaatta gattttgcat aaaaaacaga ctacataata ctgtaaaaca caacatatcc 13140
agtcactatg aatcaactac ttagatggta ttagtgacct gtagtcgacc gacagccttc 13200
caaagtgtct tcgggtgatg ctgccactt agtcgaccga cagccttcca aatgttcttc 13260
35 tcaaacggaa tcgtcgatc cagcctactc gctattgtcc tcaatgccgt attaaatcat 13320
aaaaagaaat aagaaaaaga ggtgcgagcc tcttttttgt gtgacaaaat aaaaacatct 13380
40 acctattcat atacgctagt gtcatagtcc tgaaaatcat ctgcatcaag aacaatttca 13440
caactcttat acttttctct tacaagtcgt tcggcttcat ctggattttc agcctctata 13500
cttactaaac gtgataaagt ttctgtaatt tctactgtat cgacctgcag actggctgtg 13560
45 tataaggag cctgacattt atattcccca gaacatcagg ttaatggcgt ttttgatgtc 13620
attttcgcgg tggctgagat cagccacttc ttccccgata acggagaccg gcacactggc 13680
50 catatcgggtg gtcacatgc gccagcttcc atccccgata tgcaccaccg ggtaaagttc 13740
acgggagact ttatctgaca gcagacgtgc actggccagg gggatcacca tccgtcgccc 13800
gggcgtgtca ataatatcac tctgtacatc cacaacaga cgataacggc tctctctttt 13860
55 ataggtgtaa accttaact gcatttcacc agtccctgtt ctgctcagca aaagagccgt 13920
tcatttcaat aaaccgggcy acctcagcca tcccttctg attttccgct ttccagcgtt 13980
60 cggcacgcag acgacgggct tcattctgca tggttgtgct taccagaccg gagatattga 14040
catcatatat gccttgagca actgatagct gtcgctgtca actgtcactg taatacgtg 14100

cttcatagca cacctctttt tgacatactt cgggtagtgc cgatcaacgt ctcattttcg 14160
ccaaaagtgt gcccagggct tcccgggtatc aacagggaca ccaggattta tttattctgc 14220
5 gaagtgatct tccgtcacag gtattttatc ggcgcaaagt gcgtcgggtg atgctgccaa 14280
cttagtcgac tacaggtcac taataccatc taagtagttg attcatagtg actggatatg 14340
10 ttgtgtttta cagtattatg tagtctgttt tttatgcaa atctaattta atatattgat 14400
atztatatca ttttacgttt ctctgttcagc tttctgttac aaagtgggtct cgaggaattc 14460
ggtaccccag cttggttaagg aaataattat tttctttttt ctttttagta taaaatagtt 14520
15 aagtgatgtt aattagtatg attataataa tatagttgtt ataattgtga aaaaataatt 14580
tataaatata ttgtttacat aaacaacata gtaatgtaaa aaaatatgac aagtgatgtg 14640
20 taagacgaag aagataaaag ttgagagtaa gtatattatt tttaatgaat ttgatcgaac 14700
atgtaagatg atatactagc attaatattt gttttaatca taatagtaat tctagctggg 14760
ttgatgaatt aaatatcaat gataaaatac tatagtaaaa ataagaataa ataaattaaa 14820
25 ataatatttt tttatgatta atagttttatt atataattaa atatctatac cattactaaa 14880
tatttttagtt taaaagttaa taaatatattt gttagaaatt ccaatctgct tgtaatttat 14940
30 caataaacia aatattaaat aacaagctaa agtaacaaat aatatcaaac taatagaaac 15000
agtaatctaa tgtaacaaaa cataatctaa tgctaataa acaaagcgca agatctatca 15060
ttttatatag tattattttc aatcaacatt cttattaatt tctaaataat acttgtagtt 15120
35 ttattaactt ctaaatggat tgactattaa ttaaatgaat tagtcgaaca tgaataaaca 15180
aggtaacatg atagatcatg tcattgtgtt atcattgatc ttacatttgg attgattaca 15240
40 gttgggaagc tgggttcgaa atcgataagc ttgcgctgca gttatcatca tcatcataga 15300
cacacgaaat aaagtaatca gattatcagt taaagctatg taatatttgc gccataacca 15360
atcaattaaa aaatagatca gtttaaagaa agatcaaagc tcaaaaaaat aaaaagagaa 15420
45 aagggtccta accaagaaaa tgaaggagaa aaactagaaa tttacctgca caagcttgga 15480
tcctctagac cactttgtac aagaaagctg aacgagaaac gtaaaatgat ataaatatca 15540
50 atatattaaa ttagattttg cataaaaaac agactacata atactgtaaa acacaacata 15600
tccagtcact atgaatcaac tacttagatg gtattagtga cctgtagtcg actaagttgg 15660
cagcatcacc cgacgcactt tgcgccgaat aaatacctgt gacggaagat cacttcgcag 15720
55 aataaataaa tcctgggtgtc cctgttgata ccgggaagcc ctgggccaac ttttggcgaa 15780
aatgagacgt tgatcggatt tcacaactct tatacttttc tcttacaagt cgttcgggctt 15840
60 catctggatt ttcagcctct atacttacta aacgtgataa agtttctgta atttctactg 15900
tatcgacctg cagactggct gtgtataagg gagcctgaca tttatattcc ccagaacatc 15960

aggttaatgg cgtttttgat gtcattttcg cggtaggctga gatcagccac ttcttccccg 16020
 5 ataacggaga ccggcacact ggccatatcg gtggatcatca tgcgccagct ttcatccccg 16080
 atatgcacca ccgggtaaag ttcacgggag actttatctg acagcagacg tgcactggcc 16140
 agggggatca ccatccgctc cccgggctg tcaataatat cactctgtac atccacaaac 16200
 10 agacgataac ggctctctct tttatagggtg taaaccttaa actgcatttc accagtccct 16260
 gttctcgtca gcaaaagagc cgttcatttc aataaaccgg gcgacctcag ccatcccttc 16320
 ctgattttcc gctttccagc gttcggcacg cagacgacgg gcttcattct gcatggttgt 16380
 15 gcttaccaga ccggagatat tgacatcata tatgccttga gcaactgata gctgtcgtg 16440
 tcaactgtca ctgtaatacg ctgcttcata gcacacctct ttttgacata cttctgttct 16500
 20 tgatgcagat gattttcagg actatgacac tagcgtatat gaataggtag atgtttttat 16560
 tttgtcacac aaaaaagagg ctgcacctc tttttcttat ttctttttat gatttaatac 16620
 ggcattgagg acaatagcga gtaggctgga tacgacgatt ccgtttgaga agaacatttg 16680
 25 gaaggctgtc ggtcgactaa gttggcagca tcaccgaag aacatttgga aggctgtcgg 16740
 tcgactacag gtcactaata ccatctaagt agttgattca tagtgactgg atatgttggtg 16800
 30 ttttacagta ttatgtagtc tgttttttat gcaaaatcta atttaataata ttgatattta 16860
 tatcatttta cgtttctcgt tcagcttttt tgtacaaact tgtctagagt cctgctttaa 16920
 tgagatatgc gagacgccta tgatcgcatg atatttgctt tcaattctgt tgtgcacggt 16980
 35 gtaaaaaacc tgagcatgtg tagctcagat ccttaccgcc ggtttcgggtt cattctaattg 17040
 aatataatcac ccgttactat cgtattttta tgaataatat tctccgttca atttactgat 17100
 40 tgtaccctac tacttatatg tacaatatta aaatgaaaac aatataattgt gctgaatagg 17160
 tttatagcga catctatgat agagcgccac aataacaaac aattgcgttt tattattaca 17220
 aatccaattt taaaaaagc ggagaaccg gtcaaacta aaagactgat tacataaatc 17280
 45 ttattcaaat ttcaaaaggc cccaggggct agtatctacg acacaccgag cggcgaaacta 17340
 ataacgttca ctgaagggaa ctccgggttc ccgccggcgc gcatgggtga gattccttga 17400
 50 agttgagtat tggccgtccg ctctaccgaa agttacgggc accattcaac ccggtccagc 17460
 acggcgcccg ggtaaccgac ttgctgcccc gagaattatg cagcattttt ttggtgtatg 17520
 tgggccccaa atgaagtgca ggtcaaact tgacagtac gacaaatcgt tggcggggtc 17580
 55 caggcgcaat tttgcgaca catgtcgagg ctacagcagga cctgcaggca tgcaagctag 17640
 cttactagtg atgcatattc tatagtgtca cctaaatctg c 17681

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU02/00073

A. CLASSIFICATION OF SUBJECT MATTERInt. Cl. ⁷: C12N 15/09 15/63

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

SEE ELECTRONIC DATABASE BOX BELOW

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SEE ELECTRONIC DATABASE BOX BELOW

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Medline, Chem Abs, Biosis, WPIDS

keywords: gene silencing, genetic vector, dsRNA, multiple recombination sites, recombinatorial cloning, RNAi
GenBank, EMBL: sequence IDs 13, 23-26**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|--|-----------------------|
| PX | WESLEY, S. V. et al (2001) "Construct design for efficient, effective and high-throughput gene silencing in plants" <i>The Plant Journal</i> , 27(6), 581-590. | 1-33 |
| A | MONTGOMERY, M. K. et al (1998) "Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression" <i>TIG</i> 14(7), 255-258 | |
| A | AU-A-43685/99 (Novartis AG) 2 December 1999 | |

☐ Further documents are listed in the continuation of Box C
 ☒ See patent family annex

| | |
|---|--|
| * Special categories of cited documents: | |
| "A" document defining the general state of the art which is not considered to be of particular relevance | "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "E" earlier application or patent but published on or after the international filing date | "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "O" document referring to an oral disclosure, use, exhibition or other means | "&" document member of the same patent family |
| "P" document published prior to the international filing date but later than the priority date claimed | |

| | |
|---|--|
| Date of the actual completion of the international search 15 February 2002 | Date of mailing of the international search report 04 MAR 2002 |
| Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929 | Authorized officer PHILIPPA WYRDEMAN Telephone No : (02) 6283 2554 |

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/AU02/00073

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

| Patent Document Cited in Search Report | | Patent Family Member | | | |
|---|----------|----------------------|----------|----|---------|
| AU | 43685/99 | WO | 99/61631 | BR | 9910729 |
| HU | 0102103 | PL | 344312 | EP | 1080208 |

END OF ANNEX